



Structural folding dynamics of an archetypal conformational disease using Nuclear Magnetic Resonance Spectroscopy

Ph.D. Thesis

Géraldine Rosalie Levy

Institute of Structural and Molecular Biology (ISMB)

University College London

Submitted September 2013

Supported by
wellcometrust

Declaration

I, **Géraldine Rosalie Levy**, declare that all the work presented in this thesis is the result of my work only. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. The work herein was carried out while I was a post-graduate student at University College London, Institute of Structural and Molecular biology, under the supervision of Dr John Christodoulou and Dr Bibekbrata Gooptu.

Acknowledgements

Acknowledgements have been removed from the e-thesis.

The full version can be consulted at UCL Libraries or is available upon request to the author.

Abstract

Members of the serpin (serine protease inhibitor) superfamily of proteins regulate key physiological processes through their ability to undergo major conformational transitions. In conformational diseases, native protein conformers convert to pathological species that polymerise. Structural characterization of these key transitions is challenging. Mechanistic intermediates are unstable and minimally populated in dynamic equilibria that may be perturbed by many analytical techniques. I use Nuclear Magnetic Resonance (NMR), and Circular Dichroism (CD) spectroscopy, to investigate the interrelated processes of serpin folding, misfolding and polymerisation in solution using the 45kDa prototypic serpin α_1 -antitrypsin, the recent assignment of the backbone resonances of α_1 -antitrypsin by our group, allows us to ask more sophisticated questions by a range of NMR techniques to study its structure and dynamics. In this study, I analysed early unfolding behaviour of α_1 -antitrypsin across a urea titration within what is apparently the largest two-states system yet characterised. In order to assess the dynamics of the native state, I have used hydrogen/deuterium exchange nuclear magnetic resonance spectroscopy (HDXNMR) to characterise motions on the slow (ms) timescale. I have conducted a detailed analysis of residue-specific changes in protection from exchange across a pH titration using SOFAST-HMQC. This is complemented by a detailed preliminary analysis of fast motions (ps-ns) using NMR relaxation experiments. Moreover, a *forme fruste* deficiency variant of α_1 -antitrypsin (Lys154Asn) that forms polymers recapitulating the conformer-specific neo-epitope observed in polymers that form *in vivo* was characterized in this study. Lys154Asn α_1 -antitrypsin populates an intermediate ensemble along the polymerisation pathway at physiological temperatures. Together, this study shows how the use of powerful but minimally perturbing techniques, mild disease mutants, and physiological conditions, provides novel insights into pathological conformational behaviour.

List of abbreviations

Serpin:	Serine Protease Inhibitor
FENIB:	Familial Encephalopathy with Neuronal Inclusion Body
NMR:	Nuclear Magnetic Resonance spectroscopy
SDS:	Sodium Dodecyl Sulfate
PAGE:	Polyacrylamide Gel Electrophoresis
LB:	Luria Broth
<i>E. coli</i> :	<i>Escherichia coli</i>
Cryo-EM:	Cryo electron-microscopy
CD:	Circular Dichroism
RPM:	Revolutions Per Minute
OD:	Optical Density
PEG:	Poly Ethylene Glycol
TUG:	Transverse Urea Gel
β ME:	β Mercapto-Ethanol
EDTA:	EthyleneDiamineTetracetic Acid
IPTG:	Isopropyl β -D-1-thiogalactopyranoside
ER:	Endoplasmic Reticulum
HSQC:	Heteronuclear Single Quantum Coherence
TROSY:	Transverse Relaxation-Optimized Spectroscopy
RDC:	Residual Dipolar Coupling
HMQC:	Heteronuclear Multiple quantum coherence
SOFAST:	Selective Optimized Flip Angle Short Transient
DSS:	4,4-dimethyl-4-silapentane-1-sulfonic acid
MW:	Molecular Weight
RNA:	Ribonucleic Acid

mRNA:	Messenger Ribonucleir Acid
tRNA:	Transfer Ribonucleir Acid
DNA:	Deoxyribonucleic acid
cDNA:	Complementary DNA
v/v: volume:	volume ratio
w/v: weight:	volume ratio
kDa:	kilo Daltons
$\Delta\nu_{1H}$:	Linewidth
τ_c :	Rotational correlation time
RCL:	Reactive Centre Loop
MS:	Mass Spectrometry
IM/MS:	Ion Mobility Spectrometry - Mass Spectrometry experiments
ToF:	Time of Flight
m/z:	mass:charge ratio
CID:	Collision Induced Dissociation
HDXNMR:	Hydrogen Deuterium Exchange Nuclear Magnetic Resonance Spectroscopy
HDXMS:	Hydrogen Deuterium Exchange Mass Spectrometry
FID:	Free Induction Decay
NOE:	Nuclear Overhauser Effect
FT:	Fourier Transformation
PDB:	Protein Data Base
ω_0 :	Precession rate
γ :	Gyromagnetic ratio
F_{app} :	Apparent fraction unfolded
WT:	Wild Type
ΔG :	Gibbs Free energy
ΔH :	Difference in enthalpy
T:	temperature

ΔS :	Difference in entropy
k_{ex} :	Rate of exchange
k_{int} :	Intrinsic rate
k_{op} :	Rate of opening
k_{cl} :	Rate of closing
ΔG_{app} :	Free energy between the close
GdmCl:	Guanidine Hydrochloride
K:	Equilibrium constant
F:	Folded
U:	Unfolded
I:	Intermediate
M*:	Molten Globule-like
CFTR:	Cystic Fibrosis Transmembrane Conductance Regulator
ERAD:	ER-associated degradation
EOR:	ER Overload Response
S:	Stressed
R:	Relaxed
B_0 :	Externally applied magnetic field
R_2 :	Transverse relaxation rate
τ_c :	Rotational correlation time
pET:	pTerminator Expression vector system
PCR:	Polymerase chain reaction
AmPS:	Ammonium persulphate
IM-MS:	Ion Mobility - Mass Spectrometry
DTT:	Dithiothreitol
U.V.:	Ultra Violet
DSC:	Differential Scanning Calorimetry
mAb:	Monoclonal antibody
P:	Protection factor

IEF:	Iso-Electric Focusing
SI:	Stoichiometry of Inhibition
K_{app} :	Association rate constant
PAI-1:	Plasminogen Activator Inhibitor-1
CCS:	Collision Cross Section
INEPT:	Insensitive Nuclei Enhanced by Polarization Transfer

Table of Contents

Declaration	1
Acknowledgements	2
Abstract	3
List of abbreviations	4
Chapter 1 – INTRODUCTION	12
Protein folding	13
Physical forces and principle underlying protein folding	13
1.1.1.1. Thermodynamics	14
1.1.1.2. Kinetics	16
1.1.1.3. Models of protein folding	16
1.1.1.4. Polypeptide chain behaviour and energy landscapes	18
1.1.1.5. Tools for the study of protein folding.	20
1.1.1.6. Relationship of polypeptide folding within cells and in isolation	20
1.1.1.7. Protein misfolding and conformational disease	21
1.2. Serpins and serpinopathies	23
1.2.1. α_1 -antitrypsin deficiency	24
1.2.2. Metastability and inhibitory mechanism	25
1.2.3. Alternative serpin conformations	27
1.2.4. Studies of serpin folding	29
1.2.5. Serpin polymerisation	31
1.2.5.1. Polymerisation induced by loop cleavage or peptide insertion	32
1.2.5.2. Three models of polymerisation proposed for disease mechanism and intermediate formation	34
1.3. NMR spectroscopy in studying protein folding	38
1.3.1. General principles of NMR	38
1.3.2. Study of protein structure and dynamics using NMR spectroscopy	41
1.3.3. NMR studies of high molecular weight proteins	42
1.3.4. Introduction to NMR of α_1 -antitrypsin	46
1.3.5. Multiple timescales allow structural and dynamical behaviour to be characterized	48
Chapter 2 – MATERIALS AND METHODS	54
2.1 Materials	55
2.1.1 Regents	55
2.1.2 Instruments	55
2.1.2.1. Cell purification and characterization	55
2.1.2.2. Circular Dichroism	56
2.1.2.3. Nuclear Magnetic Resonance (NMR) Spectroscopy	56
2.1.2.4. Mass Spectrometry (MS)	56
2.2 Methods	57
2.2.1 Protein recombinant expression in <i>Escherichia coli</i> (<i>E. coli</i>) and purification	57

2.2.1.1. Plasmid constructs	57
2.2.1.2 Transformation of BL-21-Gold E. coli with plasmid pQE-31 containing α_1 -antitrypsin cDNA	59
2.2.1.3 Unlabelled recombinant α_1 -antitrypsin protein production	59
2.2.1.4 Uniform incorporation of ^{15}N - and ^{13}C - ^{15}N - ^2H - isotopic labeling for NMR	60
2.2.2 Purification of α_1 -antitrypsin	60
2.2.3 Biophysical and biochemical characterization	62
2.2.3.1 Activity Assay	62
2.2.3.2 Far U.V. Circular Dichroism spectroscopy	63
2.2.3.3 Native-PAGE	63
2.2.3.4 Transverse Urea Gradient-PAGE	64
2.2.3.5 ^1H NMR	65
2.2.3.6 Native Mass Spectrometry (MS)	65
2.2.4 Hydrogen Deuterium exchange (HDXNMR)	66
2.2.4.1 Sample preparation	66
2.2.4.2 ^1H - ^{15}N SOFAST HMQC NMR for HDXNMR experiments	66
2.2.4.3. Spectral processing and NMR frequency referencing	67
2.2.4.4. Data processing and analysis	68
2.2.5. NMR studies of α_1 -antitrypsin in urea	70
2.2.5.1. Sample preparation	70
2.2.5.2 ^1H - ^{15}N -TROSY HSQC	71
2.2.6. NMR relaxation of α_1 -antitrypsin	71
2.2.6.1 Sample preparation	71
Chapter 3 – NMR STUDIES OF α_1 -ANTITRYPSIN SOLUTION BEHAVIOUR IN UREA	74
3.1. Introduction	75
3.1.1. Metastability/Z-mutant and formation of intermediate ensemble.	75
3.1.2. Equilibrium folding intermediate	76
3.1.3. Probing the folding/unfolding pathway of α_1 -antitrypsin in guanidine	78
3.1.4. Probing the folding/unfolding pathway of α_1 -antitrypsin in urea	80
3.1.5. Effect of stabilizing ‘cavity-filling’ mutations on the native state α_1 -antitrypsin	81
3.2 Results	82
3.2.1. NMR studies of α_1 -antitrypsin unfolding structural behaviour	82
3.2.2. Circular Dichroism studies of the plasma derived and recombinant α_1 -antitrypsin	92
3.2.3. Transiently populated α_1 -antitrypsin intermediates form off-pathway polymers with defined secondary structure in urea	95
3.3. Discussion	97
Chapter 4 – STRUCTURAL DYNAMICS OF THE NATIVE α_1 -ANTITRYPSIN STUDIED BY HYDROGEN DEUTERIUM EXCHANGE NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY	101
4.1. Introduction	102
4.1.1. General principle of HDXNMR	102
4.1.2. HDXMS studies of α_1 -antitrypsin	105

4.2. Results	107
4.2.1. Sample integrity	107
4.2.2. Reproducibility of HD exchange at pH 7.0.	110
4.2.3. Patterns of HDXNMR observed for α_1 -antitrypsin	113
4.3. Discussion	119
Chapter 5 – CHARACTERISATION OF THE SOLUTION BEHAVIOUR OF THE NOVEL MILD DEFICIENCY MUTANT Lys154Asn α_1 -ANTITRYPSIN	123
5.1 Introduction	124
5.1.1. Case report	124
5.1.2 Involvement of the F-helix in conformational transitions	125
5.1.2.1 Rouen-VI antithrombin	126
5.1.2.2. Leu55Pro α_1 -antichymotrypsin	126
5.1.2.3. Solution studies of F-helix remodeling	128
5.1.2.4. Potential structural effects of the Lys154Asn mutation in α_1 -antitrypsin	129
5.2. Results	130
5.2.1. Production of recombinant Lys154Asn α_1 -antitrypsin	130
5.2.2. Biochemical characterization	131
5.2.3 Activity assay and inhibitory kinetics	133
5.2.4. Thermal stability and polymerisation	135
5.2.4.1. Circular Dichroism	135
5.2.4.2. Native-PAGE of polymerisation at 37 °C and 42 °C	138
5.2.4.3. Polymerisation in mammalian cells	139
5.2.4.4. Polymerogenic intermediate formation assessed by ANS fluorescence	140
5.2.4.5. Ion Mobility-Mass Spectrometry	142
5.2.5. Assignment of the Lys154Asn α_1 -antitrypsin using NMR spectroscopy	143
5.2.6. NMR of polymerogenic intermediate ensemble at 37 °C	147
5.3 Discussion	150
Chapter 6 – DISCUSSION	154
SUPPLEMENTARY MATERIALS	166
Appendix A: Fast Timescale Dynamics from NMR relaxation experiments of α_1 -antitrypsin	167
A.1. Introduction to R_1 , R_2 , and the $^{15}\text{N}\{-^1\text{H}\}$ NOE	167
A.2. Data acquired and analysis to date:	169
A.2.2. Methods	169
Appendix B: Therapeutic strategies for blocking α_1 -Antitrypsin polymerisation: TTA1 peptide binding	173
Appendix C: Fast Timescale Dynamics from HDXNMR experiments of α_1 -antitrypsin	175
References	186

Chapter **1** - INTRODUCTION

1.1. Protein folding

Protein folding is the process by which a protein assumes its functional conformation in health. However neurodegenerative and other diseases are believed to result from an accumulation of protein aggregates formed by misfolded proteins. What factors determine the conformation of a protein? How does conformational change occur and how does this affect a protein's function and dysfunction? Can an understanding of structural biology help unveil the mechanisms of neurodegenerative diseases such as Alzheimer's and Parkinson's disease?

The number of different types of proteins in the human body approaches 100 000¹, which are engaged in promoting or controlling virtually every event upon which our lives depend. Despite the numerous studies carried out on protein folding and misfolding, the mechanism by which a protein folds remains an ongoing question. A key principle is that the reversible *in vitro* folding of a protein in a cell free system to a thermodynamically stable native state indicates that this state has the global minimum free energy of all kinetically stable structures². Christian Anfinsen and colleagues' work on ribonuclease showed that the native structure of a protein is determined by its amino acid sequence³, a finding for which Anfinsen received the Nobel Prize in Chemistry in 1972⁴. They proposed that under the environmental conditions (temperature, solvent concentration, composition) in which folding occurs, the native conformation is the unique, stable, and kinetically accessible minimum free energy state. Protein folding within a specific environment is therefore dependent upon thermodynamic and kinetic considerations.

1.1.1. Physical forces and principle underlying protein folding

A linear polypeptide chain is autonomously organized in a well-defined three-dimensional structure. In the case of a globular protein, the internal core is mostly formed by hydrophobic residues⁵. Together, it is compacted by van der Waals forces, and the globule surface is mostly formed by charged and polar side

chains⁶. The conformation of this state of condensed matter is largely determined by the flexibility of the polypeptide backbone and by the specific, consistent intermolecular interactions of the side chains.

1.1.1.1 Thermodynamics

From a thermodynamic point of view, the free energy of a protein molecule is influenced by: (i) the hydrophobic effect, (ii) the energy of hydrogen bonds, (iii) the energy of electrostatic interactions, and (iiii) the conformational entropy due to the restricted motion of the main chain and the side chains⁷.

Thermodynamic stability is defined by the Gibbs free energy (ΔG), which reflects the difference in energy between the folded native conformation (N) and the unfolded (U) state(s) of a protein. The stability of the folded conformer is a function of the fragile balance between compensating forces, being favorable hydrophobic, van der Waals, hydrogen-bonding interactions; and unfavorable conformational entropy. The folding free energy $\Delta_F G$ (of the Folded \leftrightarrow Unfolded reaction) is typically small, it is of the range 5-15 kcal.mol⁻¹ for most globular proteins⁸ (in comparison with approximately 30 to 100 kcal.mol⁻¹ for a covalent bond). It reflects the balance between enthalpy and entropy by the relation:

$$\Delta G = \Delta H - T\Delta S$$

Equation 1.1

Where ΔH is the difference in enthalpy, T is the temperature and ΔS is the difference in entropy.

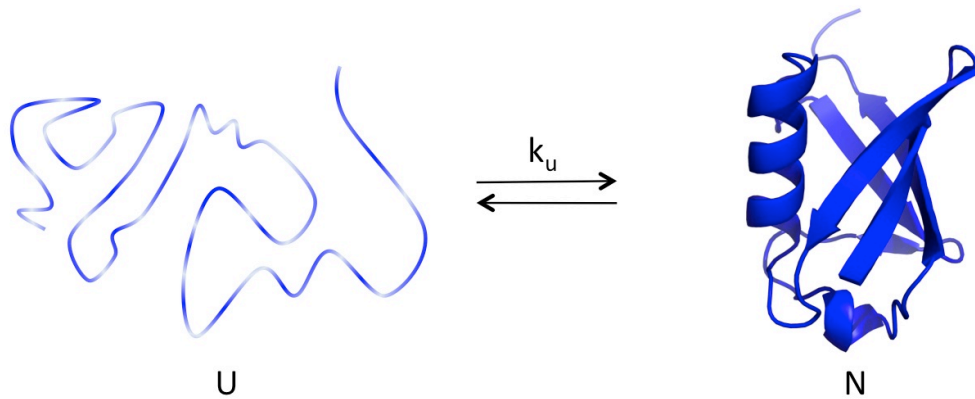


Figure 1.1. Illustration of the process of protein folding. It is the process by which a protein structure assumes its functional shape or conformation from an unfolded (U) to a fully functioning folded state (N), via an equilibrium constant of unfolding (k_u), illustrated using protein folding model ubiquitin (PDB: 1UBQ).

In the unfolded state of a protein, there is high configurational entropy. Protein folding will occur when entropy loss is offset by a gain in enthalpy, resulting from main and side-chain packing interactions. This results in a negative ΔG , and hence a favorable reaction. For any reversible reaction, from unfolded to folded, $\Delta_F G$ under standard conditions can be expressed as the relation:

$$\Delta_F G = -RT \ln K \quad \text{Equation 1.2}$$

Where $\Delta_F G$ is the free energy of folding, R is the gas constant, T the temperature in Kelvin and K_{eq} the equilibrium constant of folding where $K = [N]/[U]^n$, where $[N]$ and $[U]$ are the molar concentrations of folded and unfolded protein, respectively n denotes how many chains associate when the protein is folded (Fig.1.1.).

$$\Delta_F G = G_N - G_U \quad \text{Equation 1.3}$$

The Gibbs free energy, G , is made up the two terms enthalpy (H) and entropy (S), where $G = H - TS$.

1.1.1.2. Kinetics

The kinetics of folding within the polypeptide chain are defined by reversibility and the time needed for a protein to reach a native state from a given starting state.

The rate of folding at a given temperature equals to $k_1[U]^n$ and the rate of unfolding equals to $k_{-1}[N]$, where k_1 and k_{-1} are the rate constants for folding and unfolding, respectively. At equilibrium the rates of folding and unfolding are equal, i.e., $k_1[U]^n = k_{-1}[N]$, where the rate of folding is also defined as:

$$k_F = k_1 / k_{-1} \qquad \text{Equation 1.4}$$

The kinetics of folding within a polypeptide chain are particularly important as they limit the range of conformational space that it can explore. For most proteins this does not ultimately prevent folding to the most stable physiological state, and may assist in solving the Levinthal paradox⁹. This points out that number of possible conformations available to a given protein is vast. Thus, even a protein of a 100 residues size, would require more time than the universe has existed to explore all possible conformations (10^{10} years) in order to adopt the appropriate fold. The paradox rests on the assumption that all conformations are equally likely to be populated in the path from the unfolded to the folded states. Folding kinetics, together with thermodynamic considerations, mean this is not the case. Moreover, such kinetics can also cause a polypeptide to become trapped within local energy minima. However, intracellular environments are rich in chaperone molecules and degradative machinery that respectively facilitate escape from, or proteolysis of, these states¹⁰. Thus the most stable physiological state is generally favoured.

1.1.1.3. Models of protein folding

The concept of folding pathways motivated a vast number of experimental research aimed at finding the specific 'folding intermediates', and

also give rise to a number of models describing the folding process. The nucleation/growth model¹¹ addresses Levinthal's paradox by assuming that the rate-limiting step of the folding process is an initial nucleation event of hydrophobic collapse. Once the nucleation takes place the nuclei grow fast and the folding process rapidly completes; the secondary structure is formed, and finally these secondary structures are collided together and pack tightly together. However many systems appear to proceed via folding intermediate ensemble(s). In contrast, the 'jigsaw puzzle' or framework model¹² proposes that folding pathways are not necessarily unique and directed. Instead each protein molecule may follow a different route to the native structure. These two models of protein folding may be reconciled in the 'nucleation-condensation' model^{13; 14} in which native hydrophobic interactions stabilize otherwise variable secondary structure features to give a transition state¹⁵. The 'framework model' presents the folding process as hierarchical in which secondary structure is proposed to fold first, followed by docking of the pre-formed secondary structural units to yield the native, folded protein¹⁵.

Folding studies on a range of globular proteins ^{16; 17; 18; 19} have probed intermediate formation along the folding pathways. Equilibrium states are not necessarily identical to intermediates populated along the folding pathways. However, it was of interest that such studies indicated equilibrium intermediates possessed native-like secondary structure, which was found to form later tertiary structure, suggesting a hierarchic process.²⁰ Such intermediates are trapped within a local minima, and models involving one or more intermediate states have been generated to explain complex protein folding kinetics. These models distinguish between two intermediates: "on-pathway" and "off-pathway". The former are those considered present in the direct sequence of folding steps between the unfolded and the fully folded native state²⁰. Off-pathway intermediates are considered to represent states populated when the adopted conformation is alternative to this. They may therefore correspond to misfolded protein that must partially or completely unfold to allow correct refolding to the native conformer. In the energy landscape view described later in 1.1.1.4, there is no clear distinction between on- and off-pathway intermediates^{21; 22}.

The 'molten globule' state was proposed for the first time in 1983 on the cytochrome c model ²³, describing for the first time the existence of a third (intermediate) equilibrium state by Ohgushi and Wada. However it was not clear whether this was related to the specific folding intermediates. Nonetheless, the equilibrium intermediates were demonstrated to be remarkably similar to each other. This was defined as a thermodynamic state that conserves a native-like secondary structure content without the tightly packed protein interior (lacked a specific tertiary structure). In the case of α -lactalbumin the molten globule state appears identical to its transient folding intermediate²⁴. Molten globule states could be generated using mild denaturing conditions (such as moderate concentrations of denaturant, low pH, high temperature, and various salts)²⁵.

1.1.1.4. Polypeptide chain behaviour and energy landscapes

These findings together sparked off the need to define the relationships between free energy and protein conformation resulting in the concept of an energy landscape²⁶. For a folding system this may be defined as a mapping of the chain conformation to its internal energy, along with rules defining what configurations are accessible from a given conformation and how the protein can move between them, as illustrated in figure 1.2.

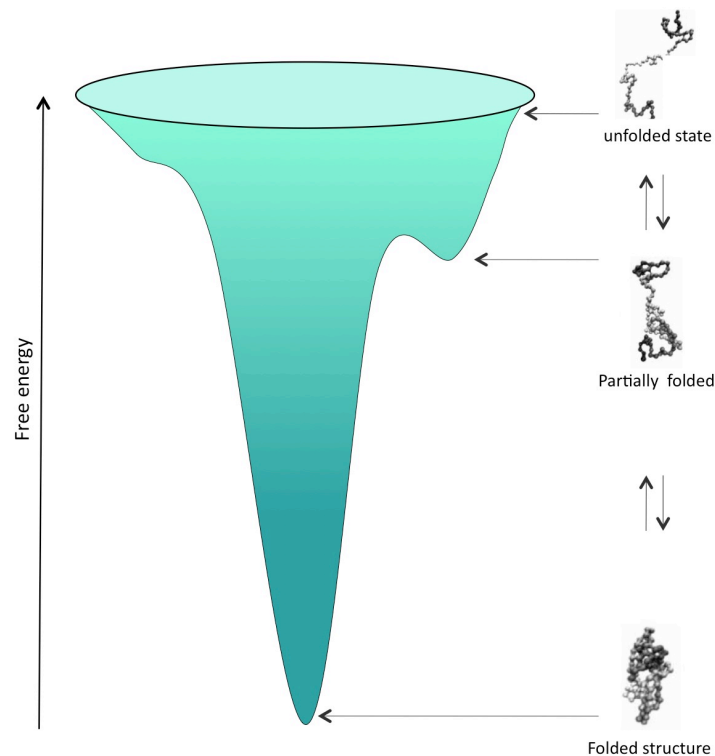


Figure 1.2. Schematic representation of the protein folding landscape. The protein energy landscape is multi-dimensional and corresponds to the interrelationships between the various configurational states of a protein. The depth represents the free energy of a given conformational state; the width is typically taken to be a measure of the configurational entropy. Different states have different energies. Each state has a transition ratio related to configurations locally connected to it. The set of states along with their energies and transition matrices fully determine the folding dynamics for the system.

Thus, the free energy of a conformational state is illustrated by the depth, while the width is typically taken to be a measure of the configurational entropy. Entropy can be defined as a measure of the number of specific ways in which a system may be arranged, often taken to be a measure of disorder, progressing towards a thermodynamic equilibrium. The enthalpy on the other hand, represents a measure of the total energy of a thermodynamic system, which takes in consideration the internal energy, as well as the volume and pressure of a given system. Recent theories of protein folding suggest that individual proteins within a large ensemble may follow a stochastic search in conformation space from the high free energy level unfolded state toward the functional fully folded native state, corresponding to the local free energy minima in their free energy surfaces, and *vice versa*^{27; 28}. The use of structural biology techniques such as Nuclear Magnetic Resonance (NMR) spectroscopy has increasingly played a

significant role in obtaining atomic-level resolution of the protein folding in solution²⁹. Indeed, NMR spectroscopy can provide detailed understanding of the nature and extent of protein dynamics on physiologically important timescales, and conditions, which can allow detailed mapping of the protein conformational energy landscape.

1.1.1.5. Tools for the study of protein folding.

In the last decade, a wide range of biochemical and biophysical methods complemented the fundamental techniques to study protein folding. Protein folding can be studied by two distinct approaches. Equilibrium studies provide information about potential folding intermediate states by characterizing stably populated ensembles across a range of folding environments. On the other hand, the relaxation approach probes the dynamics of a system as it evolves towards a new equilibrium, following rapid perturbation of the condition of its environment (pH, pressure, temperature, or solvent conditions)³⁰. Folding rates vary greatly amongst systems, with timescales between milliseconds to minutes. Classic experimental techniques to study folding kinetics are stopped-flow fluorescence spectroscopy and stopped-flow circular dichroism, that can probe the formation of secondary and tertiary structure at a millisecond timescale. Continuous-flow techniques extend the resolution time to the millisecond timescale range³¹. Using the former, studies of the initial collapse and formation of intermediates in early folding, has become possible³². The structural dynamics of these processes can be modeled *in silico* ³³.

1.1.1.6. Relationship of polypeptide folding within cells and in isolation

The energy landscape of protein folding and conformational behaviour is determined not only by its amino-acid sequence but also by the physico-chemical environment within which it exists¹⁰. The most physiologically relevant environments for protein folding are those of the intracellular folding compartments, meaning the cytoplasm or endoplasmic reticulum (ER).

Furthermore, protein folding is intrinsically linked to the macro-molecular complex named the ribosome that serves as a primary site of protein synthesis. The detailed mapping of its structure and function awarded Yonath and colleagues the Nobel Prize in Chemistry 2009³⁴. It is responsible for the translation from mRNA into all proteins of living organisms. As all proteins are synthesized, the nascent chains emerge vectorially from the ribosomal exit tunnel and into the highly crowded cellular environment to fold, often with the aid of cell machinery such as chaperones³⁵ or co-translational folding on the ribosome²⁸. Taken together the complexity of intracellular co-translational folding presents major challenges to understanding the various processes in detail. Encouraging recent progress has been made to this end by the Christodoulou group^{28; 36; 37; 38; 39} and others^{40; 41; 42; 43; 44}. Furthermore, characterizing the folding of isolated polypeptide chains in cell-free environments *in vitro*, albeit as a simplification of the physiological state, constitutes a promising avenue to aid interpretation of complex data from *in situ* studies²⁸.

1.1.1.7. Protein misfolding and conformational disease

The failure of a protein to adopt or maintain its native functional conformation, may result in a wide range of pathological disorders, referred to as 'protein misfolding diseases' or 'conformational diseases'^{45; 46; 47; 48}. Conformational disease phenotypes are related to both a loss of normal function and a gain in toxic function. It is likely that many diseases are mediated to some degree by abnormal conformational behaviour. A recent example of the importance of such mechanisms comes from cystic fibrosis⁴⁹ where small molecule stabilizers of the functional state of the transmembrane CFTR protein have been a major breakthrough in treatment^{50; 51}. However, the best-known group of conformational diseases are associated with protein deposition organized in fibrillar aggregates, e.g. amyloid fibrils^{47; 52}.

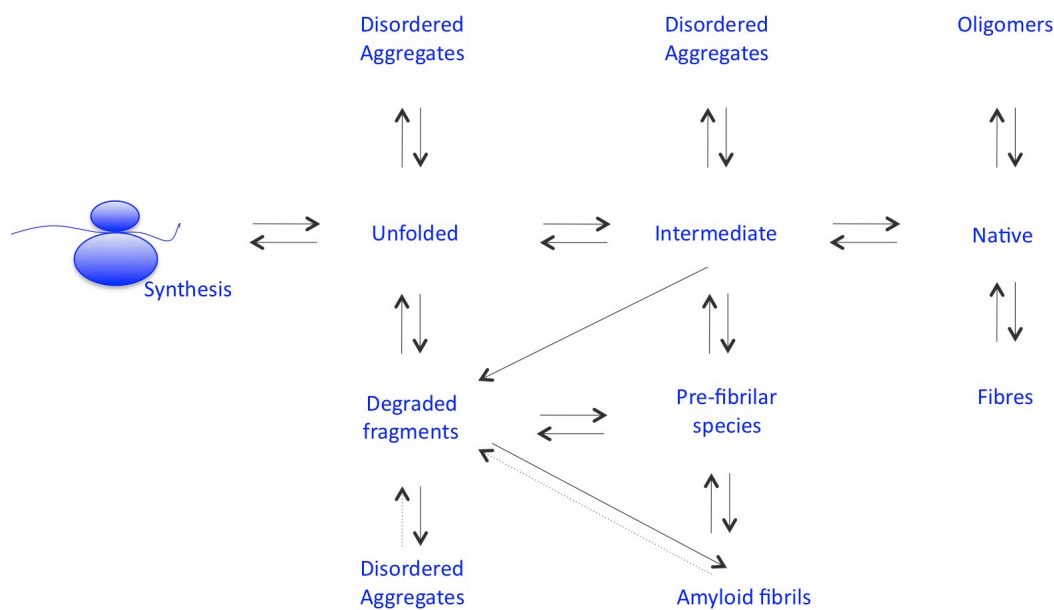


Figure 1.3. States accessible to a protein molecule. Represented here is an overview of the different states available to a molecule after it has been synthesized on the ribosome. The challenge now is to define the structures, energetic and kinetic properties of all the states of a given protein under certain conditions. (Figure adapted from the work of Dobson et al.⁵³)

The range of states available for a protein to fold is illustrated in figure 1.3. Folding into a unique three-dimensional native conformation is essential to their function, and a failure to fold into native structure generally produces inactive proteins, and in some instances, modified or toxic functionality, often characterized by the formation of aggregated proteins into amyloid fibrils. These misfolding principles are observed in a wide range of conformational diseases including Alzheimer's⁵⁴, Parkinson's^{55; 56; 57}, and Huntington's⁵⁸ diseases, and the prion encephalopathies⁵⁹. They are characterized by a 'toxic gain of function' created by the presence of accumulated protein deposition, and 'loss of function' as a result from the deficiency of activity required for their function. Conformational diseases can be inherited, usually as dominant traits, or can be induced, as in the case of prions. The vast majority of conformational diseases are, however, of unknown etiology. How cells respond to the production of these abnormal protein conformers and how these misfolded proteins cause cytotoxicity are central unanswered questions that are the subject of our study.

1.2. Serpins and serpinopathies

An archetypal model of conformational disease comes from a superfamily of proteins known as the serpins (serine protease inhibitors)⁶⁰, where disorders caused by serpin polymerisation have been termed the serpinopathies⁴⁶. The formation of polymers underlies the retention of α_1 -antitrypsin within the hepatocytes and neuroserpins within neurons to cause cirrhosis and dementia, respectively (Fig.1.4.). Similar conformational transitions were observed in other serpins (α_1 -antichymotrypsin, C1-inhibitor)⁵⁴.

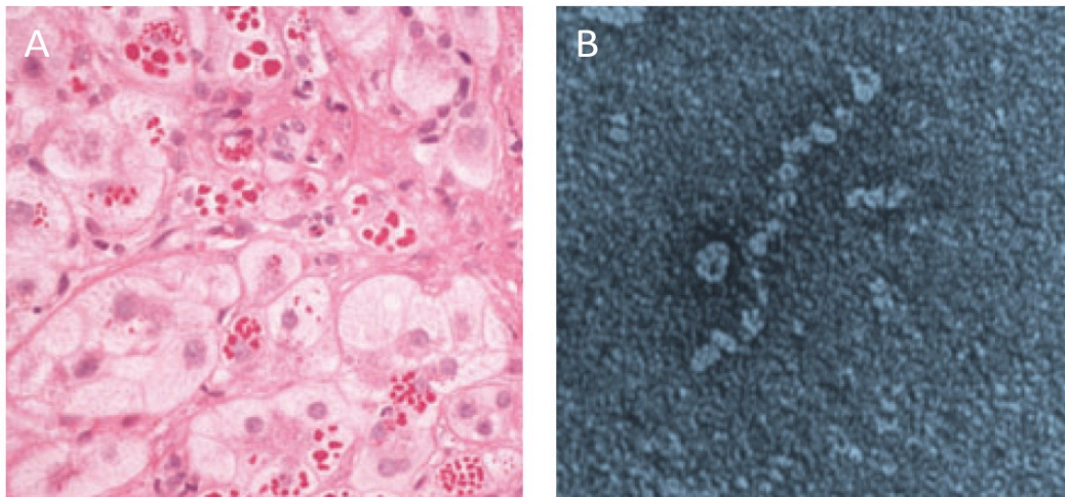


Figure 1.4. Conformational disease: Z α_1 -antitrypsin is retained within hepatocytes as intracellular inclusions. A- Intracellular inclusions are PAS positive and diastase resistant and are associated with neonatal hepatitis and hepatocellular carcinoma. B- These inclusions are composed of chains of α_1 -antitrypsin polymers shown here from the plasma of a Siyama α_1 -antitrypsin homozygote. Figure adapted from⁵⁴.

Serpins are defined by at least 30 % sequence identity to the archetypal member α_1 -antitrypsin, and a common native fold, consisting of β -sheets A, B, and C and at least 7 (typically 9) α -helices (Fig1.5.). This conformation is not the most energetically preferred state but is instead metastable⁶¹. This property underlies diverse serpin functions (protease inhibition, hormone precursors, hormone carriers and chaperones) but also predisposes to alternative stabilizing (and hence inactivating) conformational change.

Most typical serpin structures contains 9 α -helices, as in the structure of native α_1 -antitrypsin represented below, which comprises 394 amino acids. The reactive centre loop (RCL), which contains the proteinase recognition site, is an exposed, flexible stretch of ~ 20 -25 residues tethered between β -sheets A and C. The β -sheet A and the RCL domains have been highlighted in blue and red respectively, as focusing on these features indicates global changes that occur during conformational transitions.

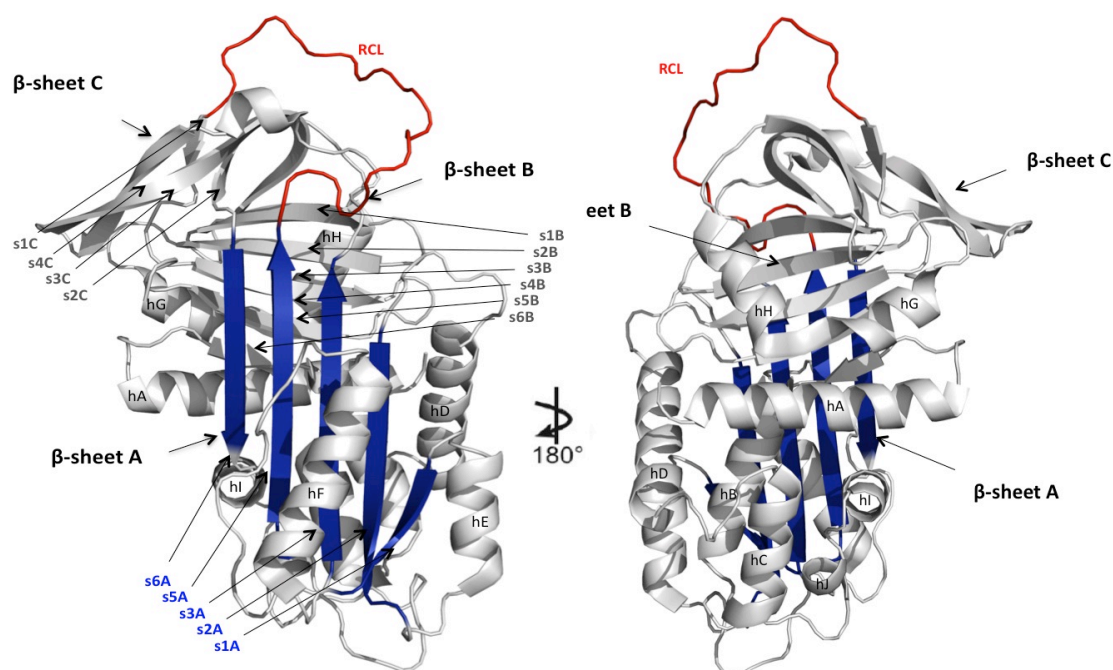


Figure 1.5. Ribbon diagram of the native state wild type α_1 -antitrypsin. The structure of α_1 -antitrypsin is composed of a lower N-terminal helical domain and an upper C-terminal β -barrel domain (when in this “classic” orientation). The main features are the RCL in red and β -sheet A in blue which are believed to undergo major conformational changes upon polymerisation, β -sheets B, and C, and 9 α -helices (A to I). (PDB: NE4)

1.2.1. α_1 -antitrypsin deficiency

α_1 -antitrypsin is the best-characterized (archetypal) member of the serpin family. Its main function is to inhibit neutrophil elastase present in the lungs and it is synthesized in the liver and circulates in the bloodstream. In the context of disease mutations, such as in the Z (Glu342Lys) variant, abnormally

folded proteins accumulate linked within polymeric chains in the ER of hepatocytes. Polymerisation is associated with gain-of-function toxicity that causes α_1 -antitrypsin deficiency associated liver disease. The accumulation of abnormally folded protein within ER causes ER stress, perturbing normal ER function. Once normal ER quality control mechanisms of autophagy^{62; 63} and ER-associated degradation (ERAD) are overwhelmed, the accumulation of abnormal proteins triggers the ER Overload Response (EOR). This is characterised by calcium-dependant NF-kappa B activation. On the other hand, the loss-of-function is the product of a lack of active circulating functional α_1 -antitrypsin mainly due to polymerisation, reduced anti-protease activity, and secretion. Together with proposed gain of function due to extracellular polymers found in lungs these effects give rise to panacinar emphysema; the hallmark of α_1 -antitrypsin deficiency-related lung disease. Polymerisation therefore seems central to disease mechanisms in α_1 -antitrypsin deficiency and the serpinopathies.

1.2.2. Metastability and inhibitory mechanism

The intrinsic dynamic potential of serpins arises because the native conformation into which they fold *in vivo* and *in vitro* is not the most energetically preferred state but is instead typically metastable⁶¹. From the native state the molecule can continue to adopt more stable conformations. This is utilized in the characteristic serpin mechanism of serine protease inhibition (Fig.1.6.). Unlike other protease inhibitors, serpins do not interact with corresponding proteases in a reversible 'lock-and-key' fashion, but function via an irreversible 'suicide-substrate' inhibitory mechanism (Fig.1.6.). The apex of the RCL of inhibitory serpins, centred on the P₁-P₁' scissile bond⁶⁴, represents an ideal substrate for the target protease in its sequence and conformation.

The initial non-covalent docking interaction (Michaelis complex) develops by the first steps of a classical serine protease mechanism. This means that P₁-P₁' peptide bond cleavage is accompanied by formation of a covalent bond

between the enzyme active site and the substrate. However this step also releases the native strain distributed around the molecule by freeing the N-terminal portion of the RCL to insert as a central antiparallel β -strand in the centre of β -sheet A (s4A). Rapid RCL insertion traps the covalently bound protease by distorting its catalytic machinery such that it cannot complete the overall hydrolysis reaction that would release a cleaved substrate⁶⁵. This final serpin-enzyme inhibitory complex is therefore physiologically irreversible. RCL insertion at the s4A site is associated with a large drop in free energy of the serpin fold and hence it is also hyperstable.

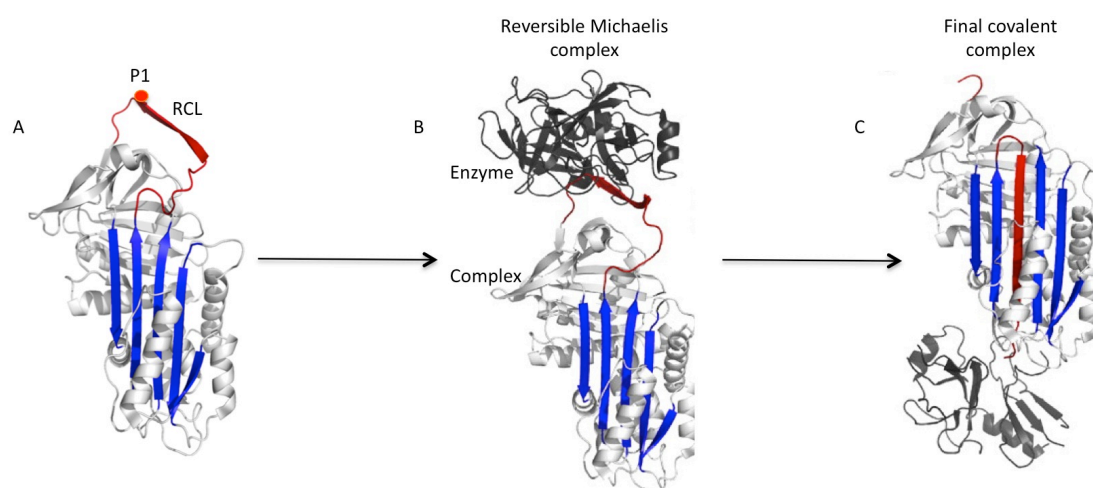


Figure 1.6. Ribbon diagram of α_1 -antitrypsin structural forms and inhibitory mechanism Inhibitory mechanism illustration from the native state can be considered as a mousetrap. A- Crystal structure of the native state of α_1 -antitrypsin where the P₁ position is represented by a red circle (PDB: 3NE4); B- Reversible Michaelis complex (PDB: 10PH). Which represents the RCL/protease interactions prior to cleavage of the P₁-P_{1'} bond and formation of the acyl-enzyme intermediate. C- Irreversible final covalent inhibitory complex (PDB: 1EZS).

The processes of serpin folding and function are therefore linked to such and extent that the inhibitory mechanism has been termed 'functional folding'⁶⁶. They can be therefore be related by a free energy landscape (Fig.1.7.), left).

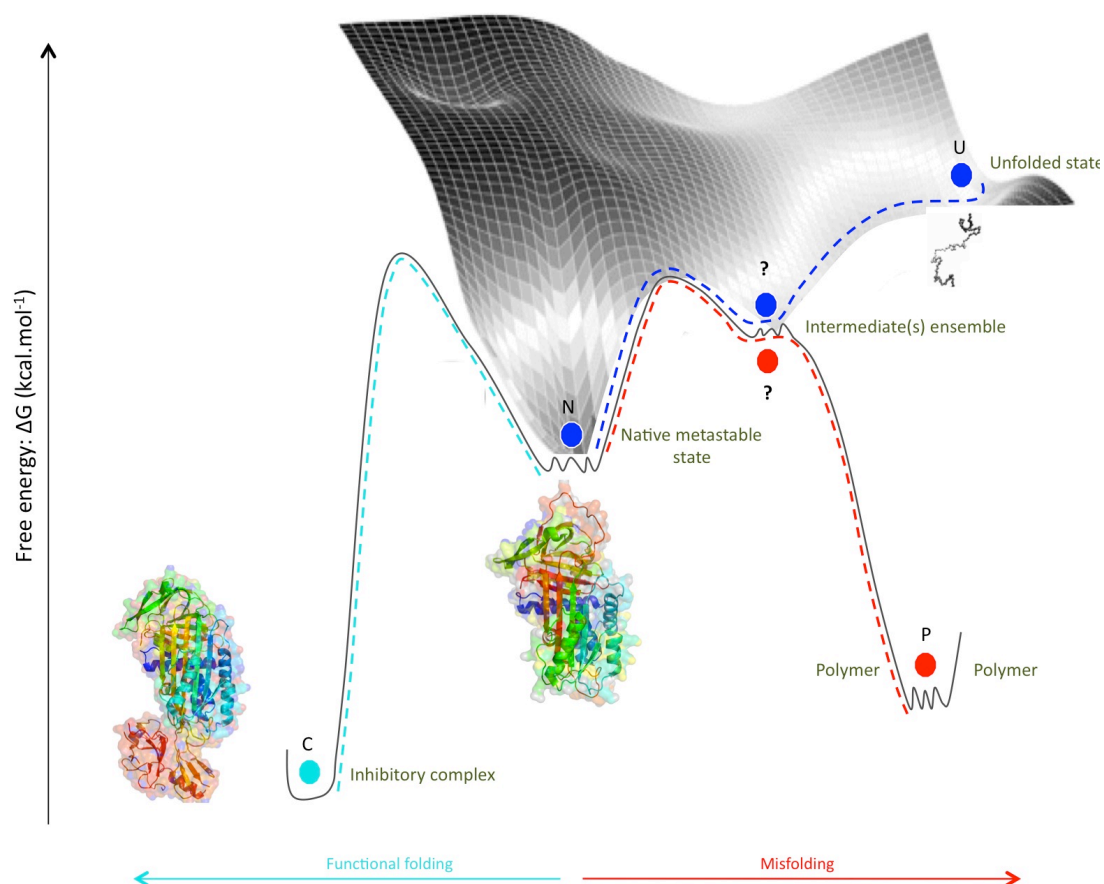


Figure 1.7. Schematic energy landscape of serpin folding and conformational change. Model of the serpin polypeptide chain folding and the relationship between folding and conformation. The surface 'funnels' the multitude of unfolded species through intermediate ensembles to the unique but dynamic native structure. From this metastable native state the molecule can adopt more stable conformations, either by functional folding required for its inhibitory function or by dysfunctional conformational change/misfolding leading to latent or polymer conformations ⁶⁷.

1.2.3 Alternative serpin conformations

The potential for stabilizing conformational change is also what renders serpins highly susceptible to the effects of disease mutations predisposing to alternative conformations outside of the inhibitory mechanism (Fig.1.7, right). Firstly, in the presence of excess or non-cognate proteases an irreversible complex may not result.

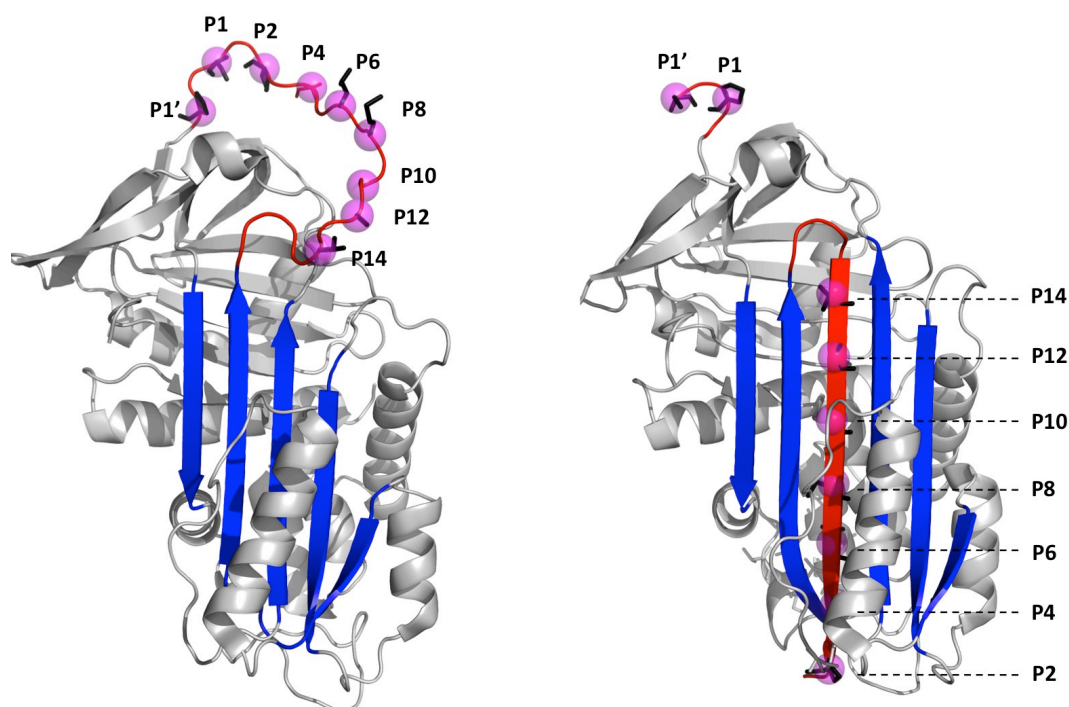


Figure 1.8. RCL P-sites upon monomeric cleavage. The RCL is highlighted in red and numbered according to the P-sites convention of Schechter and Berger, in which P_1 and $P_{1'}$ positions are either side of the site of cleavage by a protease. The β -sheet A is coloured in blue, upon transition from the native metastable conformation (left) to the cleaved (right) conformer the RCL shifts from an exposed position to an inserted position and forms an additional strand in the centre of the β -sheet A; represented for the native (PDB: 3NE4) and the cleaved conformer (PDB: 1EZK).

In these cases, failure to trap the protease in a covalent complex, leaves a cleaved serpin monomer⁶⁸. The cleaved conformation (Fig.1.8. illustrating the P-sites upon cleavage) is inactive, and demonstrates the thermodynamic hyperstability characteristic of all conformers formed from an stressed (S) to relaxed (R) transition⁶⁹, similarly to the hyperstable inhibitory complex illustrated in Fig1.7. The RCL can also insert into the β -sheet A in the absence of cleavage. This is the case of the latent species (Fig.1.9. – left) in which the RCL fully inserts into the β -sheet A⁷⁰. This latent species occurs spontaneously in PAI-1 (Plasminogen Activator Inhibitor-1), with inhibitory activity restored by denaturation and refolding, and in antithrombin⁷¹ when not bound to heparin moieties. It has also been demonstrated for α_1 -antitrypsin^{72; 73; 74}, neuroserpin⁷⁵, and α_1 -antichymotrypsin⁷⁶. The δ conformer (Fig.1.9. - middle), involves the partial insertion of the RCL in the s4A site, with the lower part of this strand

being formed through partial unwinding of the overlying F-helix, mimicking full insertion⁷⁷. Lastly, serpins can also accept an exogenous RCL analogue peptide to form protein:peptide (e.g. binary) complex, (Fig. 1.9. - right)^{78; 79; 80}.

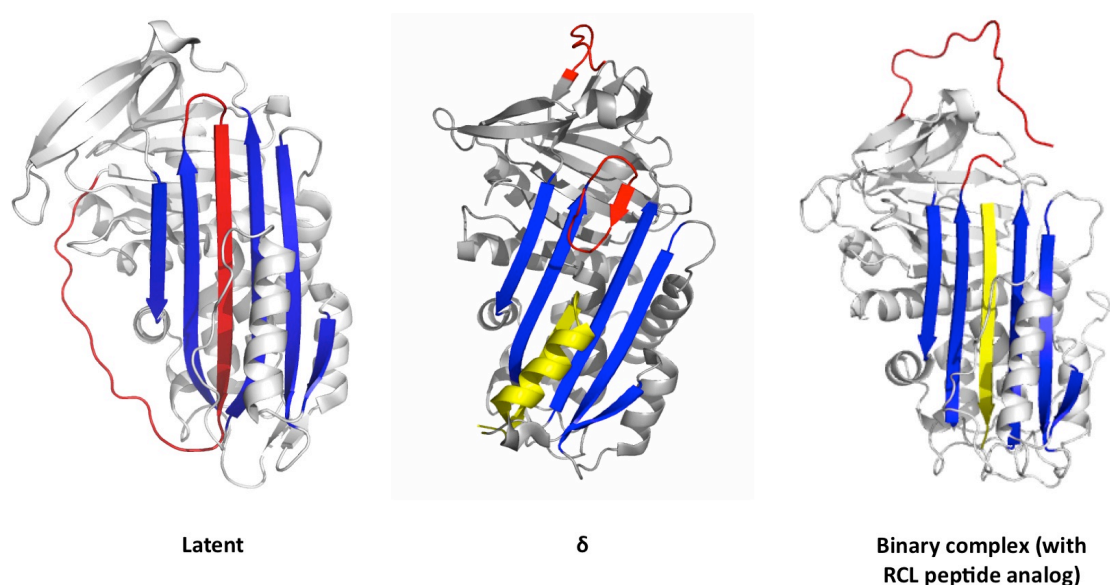


Figure 1.9. Alternative monomeric serpin conformers resulting from stabilizing conformational change. The failure to trap a substrate protease as a complex results in cleavage of the RCL (red) and its insertion into β -sheet A (blue) to generate the cleaved conformer. Serpins can also form a latent species in which the intact RCL is fully inserted into β -sheet A (left). Partial insertion of the RCL is seen in the δ conformation with the lower part of the strand (yellow) being formed by partial unwinding of the overlying F-helix (middle). Furthermore, serpins can accept an exogenous RCL analogue peptide (yellow) to form a binary complex (right).

1.2.4. Studies of serpin folding

The thermodynamic metastability of the serpin native fold is relatively unusual in protein folding and explains the propensity of serpins to undergo conformational change that has both physiological and pathological importance. However the ability to probe the folding process (and indeed other structural and dynamic behaviour) for serpins in solution is not trivial. Specific difficulties arise due to the range of conformational pathways and dynamic equilibria accessible to the fold in cell free systems. Polymerisation, and subsequent or parallel aggregation events that are less specific, present a clear challenge for data analysis, particularly when partially denatured states are studied. Moreover exogenous labeling techniques may perturb the conformational dynamics of the

polypeptide chain in solution. There is therefore a need to develop approaches that inform upon details of serpin solution behaviour without perturbing it, and where polymerisation events can be accounted for. In normal serpin folding, serpins are believed to rapidly move to a key folding intermediate to attain their native state^{81; 82; 83}. A number of studies have indicated that serpin folding intermediates have the ability to aggregate⁸⁴, it is therefore important for the polypeptide to rapidly transition to a native state. Bottomley et al. ^{81; 82; 83} have undertaken a number of serpin folding studies using increasing concentration of denaturant measured using far-UV CD^{83 81}. These data indicate that there are three species present during unfolding: the native (N), the unfolded (U) as well as an intermediate species (I)^{85; 86}. Refolding experiments showed an almost-fully reversible process ($\leq 90\%$ recovery)⁸⁷. The percentage of recovery was dependant on the time of incubation in high denaturant conditions. Together, these studies indicate that for α_1 -antitrypsin the intermediate ensemble is formed in approximately around 1.5 M guanidine hydrochloride⁸³, and around 4.0 M urea⁸¹. Tryptophan mutagenesis and fluorescence studies indicated that, during urea-induced unfolding, dynamic changes occurred around β -sheet A consistent with its expansion in the intermediate state ⁸⁸. More limited structural change was observed from a typtophan residue reporting on the environment around β -sheet B with evidence of some residual secondary structure even at 8.0 M urea. Kinetic studies of wild-type α_1 -antitrypsin in urea have further supported the involvement of 2 kinetic intermediates along the polymerisation pathway (I1 and I2)⁸⁹. I1 corresponded to an intermediate identified by Bruch et al. ^{85; 86}. However, since I2 was identified on the basis of tryptophan emission maxima data, it is important to mention that this technique may not be a robust quantitative assay ⁸⁹.

Mass Spectrometry (MS)-based methods have been developed in recent years, providing useful insights into native solution state behaviour and folding of serpins ^{90; 91; 92; 93}. One such method is Hydrogen Deuterium exchange (HDXMS). The resolution of HDXMS data is limited by the lengths of the peptide products of enzymatic treatment that can be analysed. However, these studies have provided useful insights into folding and native state behaviour of α_1 -

antitrypsin in solution. HDXMS was used to monitor region-specific folding of α_1 -antitrypsin⁹⁴. Where a key region in metastable to stable transition, β -sheet 5 A, shows protection upon folding, in contrast to the B-C barrel remains unchanged. The incorporation of the C-terminal residues into β -sheets B and C is largely complete before the center of β -sheet A begins to fold. Furthermore, findings from HDXMS of native α_1 -antitrypsin showed a highly dynamic F-helix, particularly at its C-terminal end⁹². In contrast, strands 3 and 5 of the β -sheet A, which must separate to allow insertion of the RCL, during both polymerisation and enzyme inhibition, were highly stable. Strand 1C (s1C), which anchors the distal end of the RCL is also stable, while strand 6 of the β -sheet A was highly flexible. Overall, data from HDXMS experiments on the folding of the native state demonstrated differences in the folding kinetics of different regions, particularly in key regions in the metastable to stable transition, β -sheet A, with a lag phase of 350s. The 'B-C barrel' regions shows no lag phase, and data shows the incorporation of the C-terminal residues into the β -sheet C and C is achieved before strands 3 and 5 of β -sheet A begins to fold. This was proposed as the mechanism that holds α_1 -antitrypsin in its metastable state, by avoiding more stable polymeric conformations.

1.2.5. Serpin polymerisation

In 1992, Lomas and colleagues made the crucial discovery that serpin polymerisation in the endoplasmic reticulum of liver cells most likely formed the basis of α_1 -antitrypsin deficiency⁹⁵. Stabilising conformational change can also result in the formation of serpin polymers. Polymerisation appears to be a central event in the disease mechanisms of the serpinopathies^{54; 96}. A key goal of research into these conditions is therefore to understand its structural basis, since this will imply a range of potential therapeutic targets. However, although many insights have been gained, the precise mechanism of serpin polymerisation that causes disease remains contentious with a variety of intermolecular linkage mechanisms proposed^{77; 97; 98; 99}.

Even though a lot of controversy remains around the polymerisation mechanism, a number of physical characteristics were specific to serpin polymerisation. An S-R transition is observed from the native state with concomitant loss of functional activity¹⁰⁰. Serpin polymers are generally analyzed using non-denaturing gel electrophoresis, and are characterized by ladders of oligomeric bands¹⁰¹. Using electron microscopy, polymers have been observed as 'bead-like' morphology, with substantial flexibility between subunits¹⁰². Polymerisation is concentration-dependent, and can be induced *in vitro* from the native state by heat, and chemical denaturants (e.g. urea, guanidine, low pH) and RCL cleavage between residues preceding the RCL¹⁰³.

1.2.5.1. Polymerisation induced by loop cleavage or peptide insertion

The ability of inhibitory serpins to polymerise upon RCL cleavage by non-cognate proteases N-terminal to the P₁-P₁' bond (e.g. near P₈) was first observed over two decades ago¹⁰⁴. Subsequently, crystal structures of such so-called cleaved polymers were solved^{105; 106}. Intramolecular insertion of the truncated s4A β -strand was complemented at the lower s4A site by insertion of the corresponding residues extending from the N-terminal end of strand s1C in a neighbouring molecule (Fig.1.10.).

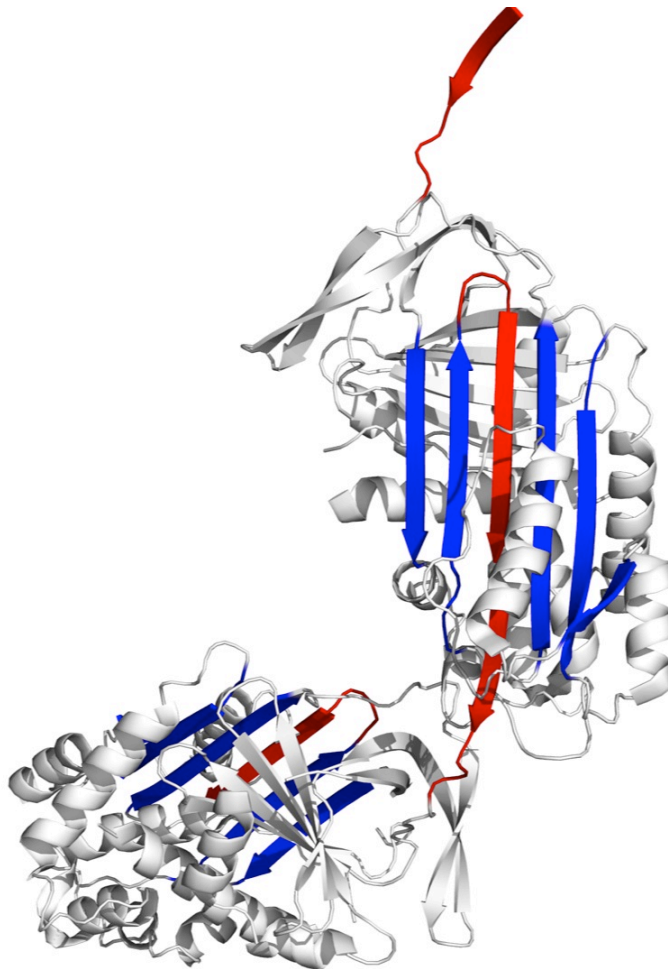


Figure 1.10. Cleaved α_1 -antitrypsin polymers at atomic resolution. A wealth of biochemical and biophysical data suggests that α_1 -antitrypsin polymers form via insertion of residues from the RCL of one molecule into the beta-sheet of another. However, this long-standing hypothesis has not been confirmed by direct structural evidence. Here is illustrated the first crystallographic evidence of a β -strand linked polymer form of α_1 -antitrypsin: the crystal structure of a cleaved α_1 -antitrypsin polymer (PDB: 1 QMB).

These data were further supported by experiments with RCL analogue peptides of different lengths¹⁰⁷. Fitton *et al.* demonstrated that incubation of the serpin antithrombin with large excess of P₁₄-P₉ exogenous peptide induced polymerisation. Conversely when peptides containing sequences corresponding to P₈-P₃ residues were used instead, polymerisation was blocked. This led to the first hypothesis that intact loop-sheet polymers could form through intermolecular insertion of the P₈-P₃ site into the β -sheet A⁹⁵. Subsequently a crystal structure of the ternary complex of antithrombin with polymerisation-inducing and blocking peptides was solved (PDB: 1 JVQ).

Together these data were therefore interpreted to support the hypothesis that partial intramolecular RCL insertion can also promote polymerisation when induced by disease mutations rather than protease activity or equivalent peptides^{104,95}.

1.2.5.2. Three models of polymerisation proposed for disease mechanism and intermediate formation

Developments over the time of these PhD studies mean that currently three models of serpin polymerisation are supported in the literature, based on high resolution snapshots obtained using X-ray crystallography^{108;109;110;111;112}. The longest established model (Fig.1.11.) can be termed the classical, loop-sheet, or single-strand model. It arises from the studies of cleaved and peptide-induced polymers. It is apparently strongly supported by the structure of the δ conformer of α_1 -antichymotrypsin since this was identified in a deficiency variant (Leu55Pro) that simultaneously favoured latent and polymeric states. Here, an unstable intermediate species is populated that is characterised by β -sheet A expansion and partial intramolecular RCL insertion. Intermolecular linkage between subunits therefore occurs via the insertion of the RCL of one constituent subunit as a central strand within the β -sheet A of its neighbour in the polymer chain. This model was most recently supported by studies of the effects of substitution of Asp residues into the reactive loop upon the polymerisation of α_1 -antitrypsin¹¹³. These showed that the presence of the charged Asp residues at P₄ and P₆ impeded polymerisation at the intermolecular linkage step. In contrast, mutation at P₈ and P₁₀ instead impaired intermediate formation.

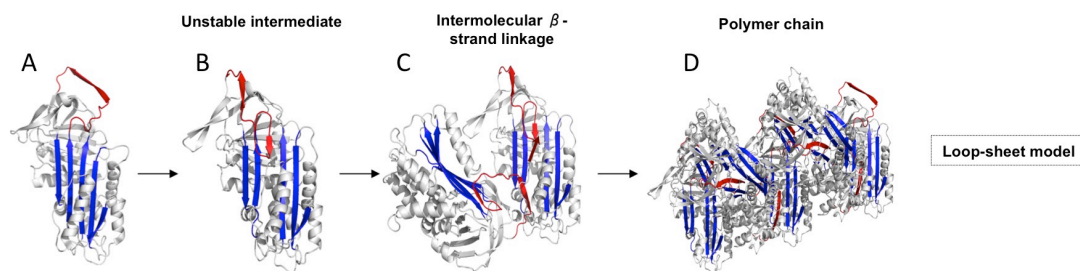


Figure 1.11. Ribbon diagram of first model of polymerisation of α_1 -antitrypsin. Model of polymerisation in which conformational change from the native conformation (A) involves the insertion of the RCL of the target proteinase as an extra strand into β -sheet A (blue) similar to what is observed in the inhibition mechanism, and inactivation of the catalytic triad of the protease. Point mutations subvert this mechanism and cause aberrant conformational transitions (B) and the formation of polymers (C and D).⁹⁶

A second model was proposed in 2008, based on the study on a closed dimer of antithrombin under low pH and low salt conditions¹¹⁴. This new model demonstrated an intermolecular β -hairpin swap involving the RCL and strand 5 of β -sheet A (s5A) comprising 50 residues, rather than single β -strand insertion. This polymerisation mechanism caused expansion of β -sheet A to allow insertion of the β -hairpin of a neighbouring molecule (Fig.1.12.). This study suggested that if considerably more structure is unfolded relative to the native fold (helix I and linker to s5A), an extended polymer could be propagated. This crystal structure of an intact serpin dimer put forward the idea that much larger “domain swaps” are possible and would also lead to hyperstable linkage between serpins monomers.

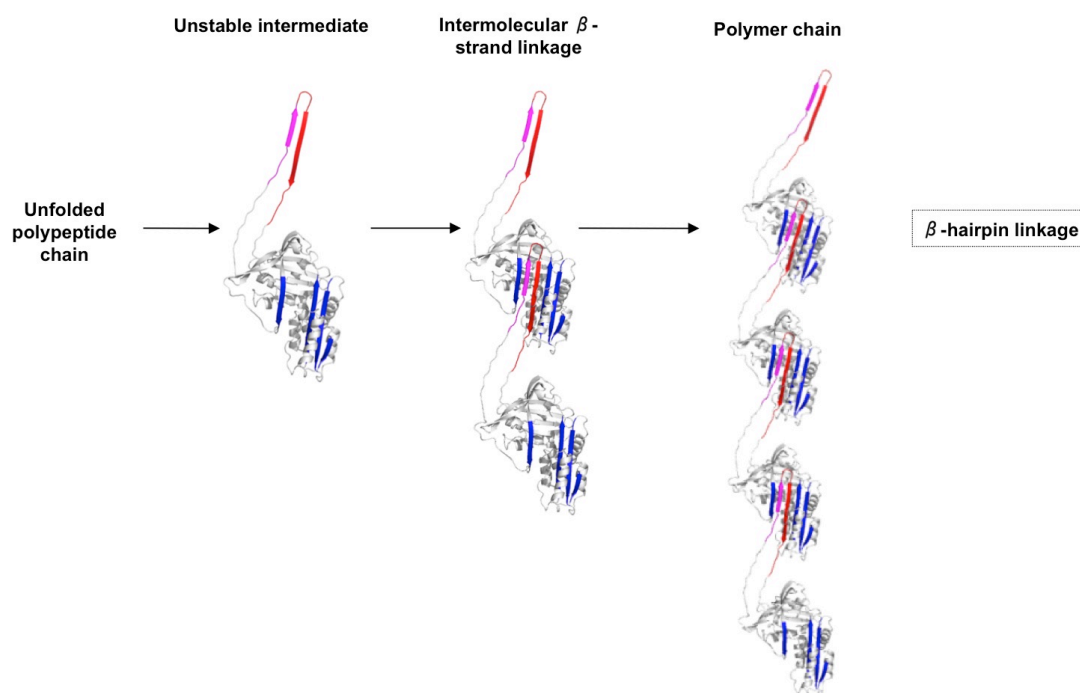


Figure 1.12. Ribbon diagram of second model of polymerisation of α_1 -antitrypsin. Model of polymerisation in which conformational change from the native conformation involve the insertion of the RCL of the target proteinase and the strand 3 of the β -sheet A as a β -hairpin on a closed dimer of antithrombin¹¹⁴.

More recently still a third model of polymerisation has been proposed (Fig.1.13). This is based on the crystal structure of a closed trimeric form of α_1 -antitrypsin. This is recognized by the 2C1 monoclonal antibody that has high conformational specificity for and between α_1 -antitrypsin polymers. This recognizes polymers observed in disease or when formed by heating, but not those formed under the conditions in which the data supporting the β -hairpin model were collected. The new structure revealed a polymeric linkage mediated by domain swapping of the C-terminal 34 residues spanning β -strands s1C, s4B and s5B¹¹⁵. In this mechanism, the stabilized expansion of β -sheet A that characterizes S-R transitions occurs by full intramolecular insertion of the RCL.

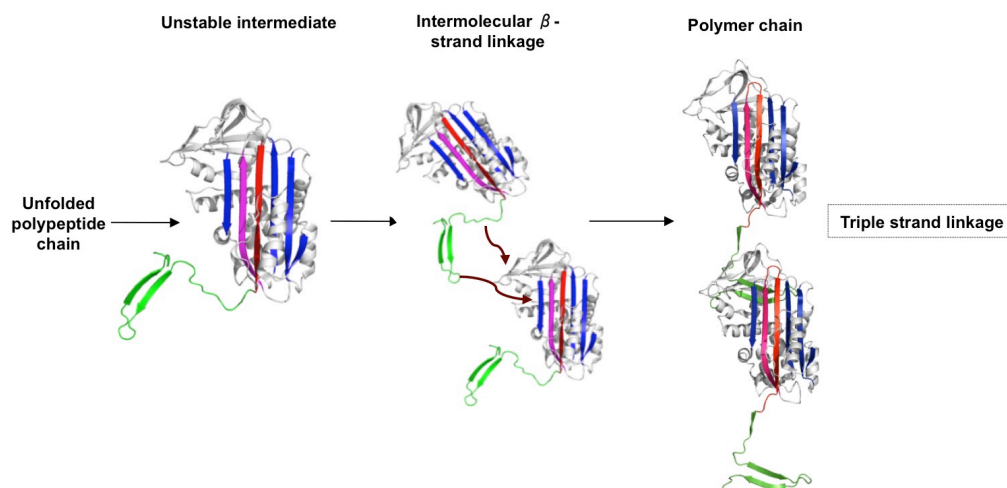


Figure 1.13. Ribbon diagram of the third model of polymerisation of α_1 -antitrypsin. The unfolded state (U) misfolds to an intermediate (I) in which the RCL (red) is fully inserted as the fourth strand of β -sheet A, and the C-terminal elements strands 1C and 4 and 5B, (green). This can attain the global free-energy minimum by forming polymers (P) via a C-terminal domain-swap mechanism as seen in the crystal structure of a self-terminating α_1 -antitrypsin trimer.

The different polymerisation models share some conceptual and structural features, but must proceed through distinct intermediate states. Moreover, since all the models of polymerisation involve formation of relatively unfolded intermediate states, it is crucial to understand the structure of those folding transitions to elucidate the polymerisation process.

It becomes clear that despite their potential importance to develop new therapeutic strategies, data available about the structure and the dynamics of the one or more intermediate state(s) along the folding pathway is limited, and unclear.

1.3. NMR spectroscopy in studying protein folding

1.3.1. General principles of NMR

Nuclear Magnetic Resonance or NMR was first described and measured in molecular beams by Isidor Rabi in 1938 (awarded Nobel Prize in Physics 1944)¹¹⁶. Over the next 50 years, NMR has developed into a powerful technique providing detailed atomic information in solution and is now a major method for studying biomacromolecules in solution. Providing high-resolution structure and dynamics. The application of NMR to biomolecules was pioneered by Wüthrich *et al.*^{117; 118} (awarded a Nobel Prize in Chemistry 2002).

The phenomenon of NMR occurs when nuclei of specific atoms are immersed into a static magnetic field and exposed to a second oscillating magnetic field. Nuclei that possess that property termed 'spin' will experience respond by demonstrating a phenomenon known as the Larmor precession. The general term of precession can be defined as a change of orientation of the rotating axis of the rotating body. It can be defined as a change in the first Euler angle, leading to the rotation of a body that rotates itself around another axis. In physics, there are two types of precession: torque-free and torque-induced. Using NMR, the Larmor precession of the nuclear spins occurs under a magnetic field that exerts a torque on the magnetic moment.

$$\vec{\Gamma} = \vec{\mu} \times \vec{B} = \gamma \vec{J} \times \vec{B} \quad \text{Equation 1.5.}$$

Where $\vec{\Gamma}$ is the torque, $\vec{\mu}$ is the magnetic dipole moment, \vec{J} is the angular momentum vector, \vec{B} is the external magnetic field, \times symbolises the cross product, and γ is the gyromagnetic moment.

The magnetic property specific to each spin can be utilized to yield chemical information. In quantum mechanical terms, the Larmor frequency or the precession rate (ω) of the nuclear magnetic moment of a nucleus is aligned with or against an externally applied magnetic field of magnitude B, and is function of the gyromagnetic ratio: γ , by the relation:

$$\omega = -\gamma B \quad \text{Equation 1.6.}$$

Each nucleus has a characteristic γ value that is defined as a constant of proportionality between the nuclear angular momentum and magnetic moment. Precession of a spin generates an electric field with ω frequency. Since γ relates the magnetic moment μ and the spin number I for a specific nucleus, the rotational axis of the precessing spin is characterized by:

$$\gamma = 2\pi\mu/\hbar I \quad \text{Equation 1.7.}$$

Where μ is the magnetic dipole moment, \hbar is the reduced Planck constant or Dirac constant and corresponds to $\hbar = h/2\pi$, I is the spin number.

The gyromagnetic ratio also corresponds to:

$$\gamma = -eg/2m \quad \text{Equation 1.8.}$$

Where e is the charge, g is the g -factor, and m is the mass.

Therefore the Larmor frequency becomes:

$$\omega = egB/2m \quad \text{Equation 1.9.}$$

Where ω is the Larmor frequency, m is the mass, e is the charge, and B is the applied magnetic field.

By irradiating the protein sample with radio waves (in the MHz frequency range), the proton will absorb the energy and progress to a less favourable energy state. Electromagnetic radiation is re-emitted as it returns to the more favourable state. The observed resonance relates to the tendency of the nucleus to switch energy states, the frequencies of the applied radiation and the precession. The specific resonance frequency is therefore a function of the strength of the applied magnetic field and the magnetic properties of the isotope

of the atoms being observed. In NMR experiments, the electromagnetic radiation applied must be of the correct frequency to allow resonant absorption by the nuclear spin, equal to the Larmor precession rate), to match the energy difference between nuclear spin level in a constant magnetic field of the appropriate strength.

The absorbing photon has energy equal to:

$$E = h\nu_0 \quad \text{Equation 1.10.}$$

Where h is the Planck constant and ν_0 corresponds to the resonance radiofrequency (equivalent to Larmor precession frequency).

Therefore, the magnetic resonance absorption will occur only when

$$\Delta E = h\nu_0 \quad \text{Equation 1.11.}$$

$$\text{when } \nu_0 = \gamma B_0 / (2\pi)$$

The NMR properties of a nucleus, such as the energy difference between the orientations and therefore the frequency, at which a residue absorbs energy, also depends strongly upon its chemical environment. Therefore, different atomic nuclei within a molecule resonate at different frequencies for similar given magnetic field. Magnetic nuclei are also affected by each other, through chemical bonds or via their spatial distance. This scalar and dipolar coupling, respectively, can be used to assign resonances to a specific nuclei within the structure, and can be used to derive constraints for their separating distance and orientation.

1.3.2. Study of protein structure and dynamics using NMR spectroscopy

NMR study can provide a detailed understanding of the structure of proteins in solution, permitting the study of intermediate species in solution^{119; 120; 121; 122}, it can also provide information about the folding of a protein^{119; 123; 124} in a residue-specific manner. Structural data describing intermediate state(s) along a folding pathway are limited through the intensive studies required to obtain relevant information but nonetheless several intermediate states have been reasonably well characterized^{125; 126; 127}. Such understanding of the structural properties of key intermediates is likely to be increasingly important in the development of target structures for therapeutic interventions. Using NMR spectroscopy, three complementary strategies can be applicable for obtaining structural information about intermediate species. The first involves characterization of species generated transiently during refolding of denatured proteins, either in real time or by means of trapping experiments¹¹⁹. The second involves the study of those partially folded states, such as the increasingly recognized molten globule state²⁹, which are stable under equilibrium conditions. The third strategy involves the design and study of models of folding intermediates. Important key NMR experiments in such studies¹²⁸ are measurements of dynamics parameters, in particular relaxation measurements and modern derivatives such as dispersion measurements, and also hydrogen-deuterium exchange (HDXNMR) experiments. HDXNMR was a major part of the work undertaken in this study and is described further below. As for relaxation experiments, these are briefly introduced in the Appendix A.

HDXNMR provides a key route to understanding protein stability, dynamics, global and sub-global fluctuations, and towards the characterization of intermediate states. Lysozyme is probably the best studied example using this method and due to early knowledge of its amino-acid sequence and three-dimensional structure, it became a central model molecule for *in vitro* study of the folding of secondary and tertiary structural element. The work of Dobson and colleagues¹²⁹ on lysozyme using HDXNMR (details are explored in Chapter IV), was a major breakthrough for the structural study of intermediate states.

The study allowed identification of hydrogen-bonding patterns in early transient folding intermediates. The pattern of amide protection within the different helices of the lysozyme was found to vary significantly. These studies, combined with mass spectrometry studies, revealed the transient unfolding of the alpha domain of lysozyme as a key intermediate in the misfolding of lysozyme.

1.3.3. NMR studies of high molecular weight proteins

Traditionally, NMR studies were restricted to those involving relatively small systems (ranging from 10 to 30 kDa in size). This range, is one where the majority of resonances can be observed by NMR and this 'limitation' is mainly caused by the relaxation processes, which are mediated by bipolar interactions and chemical shift anisotropy¹³⁰. These are highly dependent on the size of the proteins being under study; rates will increase accordingly to the size of the protein rendering the study of large molecular weight proteins using NMR highly challenging process. However, with recent developments in the NMR methodology, structure determination and dynamics studies have now been rendered feasible for molecules of large molecular weight. Indeed, the introduction of Transverse Relaxation-Optimized Spectroscopy (TROSY)¹³¹, Residual Dipolar Coupling¹³¹ (RDC) and labeling strategies¹³² has opened new avenues for NMR of high molecular weight complexes. During NMR experiments, the relaxation of the magnetization is characterized by a loss of NMR signal, the main factor responsible for this is the transverse relaxation¹²⁸. The rapid transverse relaxation rate (R_2 , characterized by the time constant T_2) and consequent loss of signal induced with increasing molecular weight increases in proportion to the mass of the molecule. This is due to the increasingly slow tumbling time, characterized by the rotational correlation time (τ_c) of the large molecules¹³². The T_2 is inversely proportional to the molecular tumbling time:

$$\tau_c = 1/T_2 \quad \text{Equation 1.12.}$$

The slower the tumbling time the broader the linewidth characterized by a loss of signal lifetime. Therefore, improving the linewidth of the observed

signal is crucial to obtaining the necessary improved sensitivity. The linewidth of the observed frequency is characterized by $\Delta\nu_{1/2}$ and depends on the ability of the excited spin to relax through T_2 mechanisms. The slower the relaxation, the sharper the linewidth of the observed frequency, characterized by the relation:

$$\Delta\nu_{1/2} = 1/(\pi.T_2). \quad \text{Equation 1.13.}$$

As a result of fast relaxation induced by large molecular weight molecules, the Free Induction Decay (FID) loses in resolution of the observed frequency after Fourier transformation. This increase in linewidth has very important consequences for high resolution NMR. It can lead to insolvable overlap issues, a significant loss in the signal to noise ratio. Implementation of a relaxation optimized¹³¹ element in the pulse sequence during the time course of the NMR experiment¹³³ has been developed to reduce the extent of loss of signal. Furthermore, labelling strategies also participate in alleviating the signal loss; using highly deuterated background and selective ^1H , ^{15}N and ^{13}C labelled residues in an otherwise ^{14}N , ^{12}C unlabeled background reduces the significant dipolar coupling arising from the high density of protonation within a polypeptide chain. This strategy therefore allows suppression of a significant contribution of the transverse relaxation from dipole-dipole interactions and scalar coupling. These developments, termed Transverse Relaxation Optimized Spectroscopy (TROSY)^{131; 134} and the methyl analogue, the methyl-TROSY HMQC¹³⁵ type of experiments have opened avenues to study biomolecules with large molecular masses. TROSY-based spectroscopy is optimal for observation of the ^{13}C methyl groups because of the increase in the recorded intensity, which arises from three equivalent protons, and the higher mobility of the side-chains compared to the backbone. Selective ^{13}C , ^1H labelling of methyl groups in a ^{12}C , ^2H background can be realized by using labeled-leucine, isoleucine, and valine precursors. Astonishing examples of structural and dynamical NMR studies of high molecular weight biological system are studies of the 670 kDa 20S proteasome of the GroEL-ES and of the SecA machinery by Kay and colleagues¹³⁶. This strategy, helped preserving the resulting NMR signal lifetime, revealing functionally important motions and interactions¹³⁷. These studies provided

detailed insight to the possible use of NMR to investigate aspects of supra-molecular structures over an order of magnitude larger than those routinely studied using methodology that is generally applicable.

1.3.3.1. HSQC

Heteronuclear Single Quantum Coherence (HSQC) or Heteronuclear Single Quantum Correlation experiments are frequently used in NMR spectroscopy for the study of large molecular weight proteins. These experiments result in a two-dimensional spectrum with one ^1H axis and the other heteronucleus axis corresponding to the isotope used for the experiments, generally ^{13}C or ^{15}N . It is designed to involve the transfer of magnetization on the proton to the second nucleus via an INEPT function (Insensitive Nuclei Enhanced by Polarization Transfer). A time delay (t_1) is applied before the magnetization is transferred back to the proton, through a so-called retro-INEPT process. Following these steps the signal will be detected in the direct dimension for protons and indirect dimension for ^{15}N or ^{13}C . This technique increases the signal to noise ratio by the use of inverse NMR experiments in which the magnetization is transferred from protons to the hetero nucleus (Fig.1.14.).

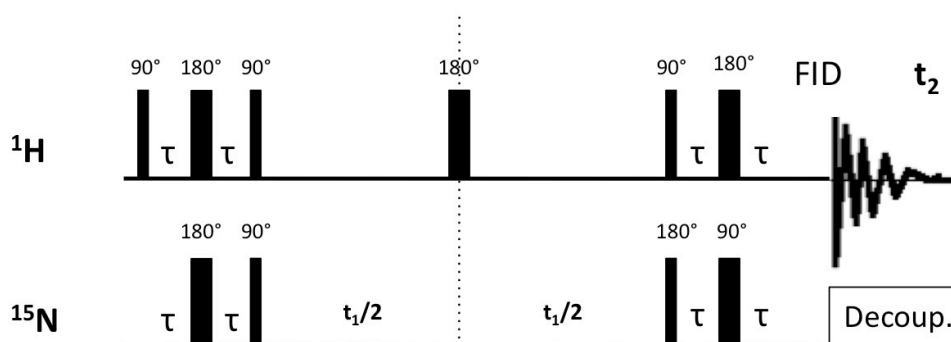


Figure 1.14. Standard HSQC Heteronuclear Single Quantum Correlation pulse sequence. The overall pulse sequence can be summarised as follows: INEPT + echo + INEPT. The first INEPT is to improve the sensitivity of the low magnetogyric ratio nuclei (^{15}N , ^{13}C). During echo, t_1 , the coupling is refocused. The retro-INEPT is to perform inverse detection.

1.3.3.2. TROSY

TROSY or Transverse Relaxation Optimized spectroscopy has been designed to be particularly suitable to large molecular weight proteins. As described using the equation 1.12, the transverse relaxation (t_2) time is directly proportionate to the tumbling time itself directly proportionate to the size of the protein. Therefore the larger the protein the greater the decay of NMR signal leading to line broadening, and a poor resolution. Typically a ^1H - ^{15}N -HSQC experiment is run with decoupling, the splitting of the resonant peak is removed during the pulse sequence. This results in a single peak that is the average chemical shift of the four in the multiplet. Each of the different multiplet components have different width, due to various interactions between relaxation mechanisms, hence relaxing at different rates. The TROSY pulse sequence is designed to select the multiplet component for which the relaxation mechanisms have almost cancelled each other out, leading to a single, sharp peak. For a large molecular weight protein this has the advantage of both increasing the sensitivity (sharp peaks) and the resolution (allowing to alleviate overlapping peaks issues occurring with large molecular weight proteins). Compared with the HSQC sequence illustrated in Fig.1.14. TROSY pulse sequence is lacking a 180° pulse applied to the proton during t_1 phase, furthermore, it lacks the decoupling of the ^{15}N nucleus. This leads to detection of a multiplet separated by J_{HN} (J-coupling). The spectra are subsequently simplified by suppression of all but the wanted peak.

1.3.3.3. SOFAST-HMQC

Two-dimensional ^1H - ^{15}N band-selective optimized-flip-angle short-transient (SOFAST)-heteronuclear multiple-quantum (HMQC) NMR experiments are optimized to allow fast and sensitive data collection. Their main component is the use of a shaped ^1H pulse that manipulates only a subset of proton spins in order to enhance the spin-lattice relaxation (T_1) via dipolar interactions with the unperturbed proton spins (Fig.1.15). The second important feature of SOFAST-HMQC experiments is the use of Ernst angle excitation to achieve maximal

sensitivity allowing high repetition of the pulse sequence. The recycle delay is therefore generally reduced to a few milliseconds and the flip angle adjusted to allow the highest overall sensitivity. SOFAST-HMQC experiments are particularly useful for real-time experiments such as HDX-NMR (described in Chapter IV) as it allows to significantly increase the sensitivity and fast repetition rates.

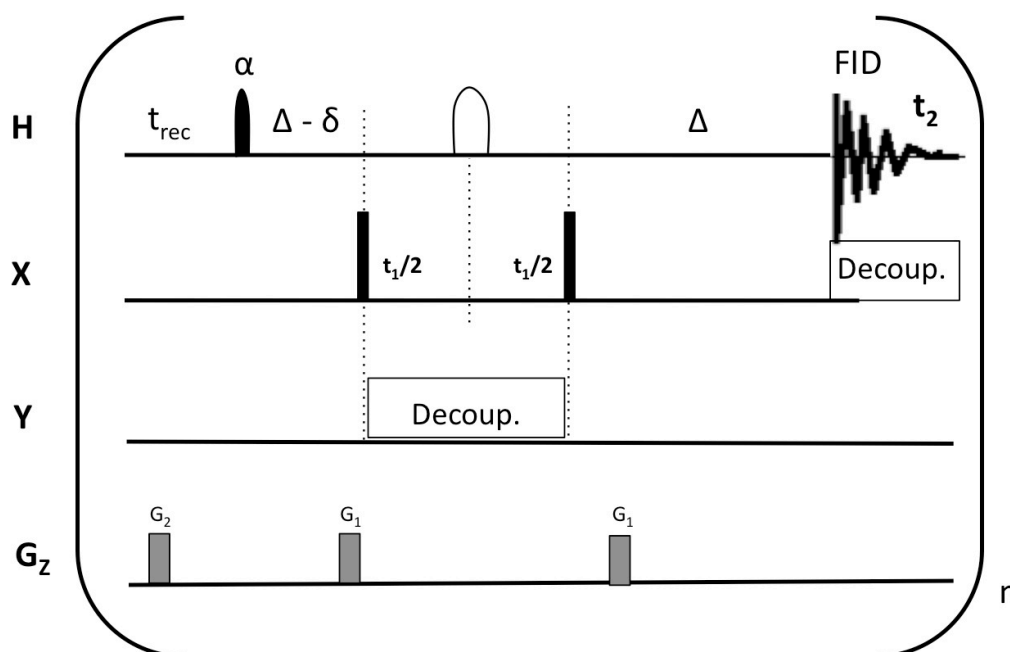


Figure 1.15. Standard SOFAST-HMQC pulse sequence. ^1H -X (X = ^{13}C or ^{15}N) correlation spectra. Filled and open pulse symbols indicate 90° and 180° rf pulses, except for the ^1H excitation pulse applied with flip angle α . The variable-flip-angle pulse has a polychromatic PC9 shape and band-selective ^1H refocusing is realized using either REBURP or r-SNOB. The transfer delay Δ is set to $1/(2J_{\text{HX}})$, and there is a recycle delay between scans. The delay δ accounts for spin evolution during the PC9 pulse, and has to be adjusted prior to data acquisition to yield pure-phase spectra in the ^1H dimension. To avoid extensive first order phase corrections in the t_1 dimension of experiments, for t_1 delays shorter than the REBURP pulse length the 90° - t_1 - 90° pulse sequence element on the ^{15}N channel is applied simultaneously to the REBURP pulse on the ^1H channel. Figure adapted from the work of Schanda et al.¹³⁸.

1.3.4. Introduction to NMR of α_1 -antitrypsin

While, as discussed the application of NMR to large systems is now becoming more widespread, it still remains challenging to apply NMR on an all-residue basis to globular proteins as large as serpins (45 kDa). However, the successful assignment of the NMR spectrum, when combined with

complementary high-resolution X-ray crystallography data, has provided a powerful tool for structural and dynamical studies of serpins folding, misfolding, and polymerisation, as well as enabling drug binding studies. The first study of serpins using NMR spectroscopy was undertaken by Gettins and colleagues¹³⁹. Using two-dimensional heteronuclear single quantum correlation 2D NMR spectra of serpins combined with specifically ¹⁵N-labelled alanine residues, the vastly simplified resulting 2D spectrum showed low resonance overlapping, and led to the assignment of a few of the observed resonances¹³⁹. Recently, the assignment of the backbone resonances of the full-length wild-type α_1 -antitrypsin using multidimensional heteronuclear NMR methods, was achieved within our group¹⁴⁰. This is one of the largest globular proteins assigned to date (Fig.1.16.). This (close to) complete backbone assignment provided the starting point for a detailed solution state characterization of α_1 -antitrypsin under a range of conditions to explore structure, dynamics, aggregation and binding interactions of the protein.

with atomic resolution. These motions vary across various timescales as illustrated alongside the range of NMR techniques available to measure these motions. Indeed, all the nuclei within the molecule can act as hidden observers of the conformational fluctuation that accompany biological activity. A number of recent NMR techniques have been developed to quantify motions within a molecule. These can take place within a wide range of timescale from the ps, to days as illustrated in figure 1.17.

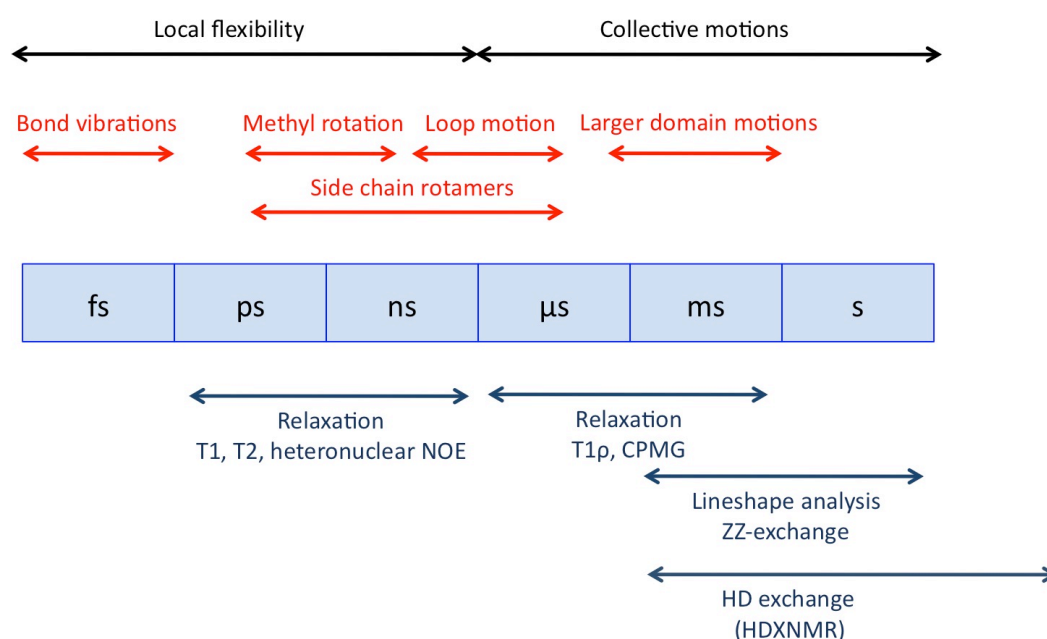


Figure 1.17. NMR techniques to measure conformational fluctuations on a range of timescales. A number of recent NMR techniques have been developed to quantify motions within a molecule. These, can take place within a wide range of timescale from the ps, to days, where the fast motions can be measured using Relaxation T1, T2, heteronuclear NOE (ps-ns), and slower motions (ms-s) can be measured by hydrogen deuterium exchange (HDXNMR) experiments.

Generally divided in two regimes, the slower timescale dynamics (μs-ms), are particularly useful to characterise a two-state process (e.g. open-closed conformation of a protein, folded/unfolded). These NMR techniques can measure exchange rates (kinetics), populations (thermodynamics) and chemical shifts (structural information) as illustrated in figure 1.18. To enumerate them, lineshape analysis, ZZ-exchange (multidimensional NMR chemical shift correlation experiments used to measure the kinetics of interconversion

between different molecular states. Its pulse sequence contains an additional mixing delay between the indirect and direct detection periods) and HD-exchange methods are mostly used for monitoring slow timescale motions. The mechanisms underlying hydrogen-deuterium exchange technique will be described in further detail in Chapter IV.

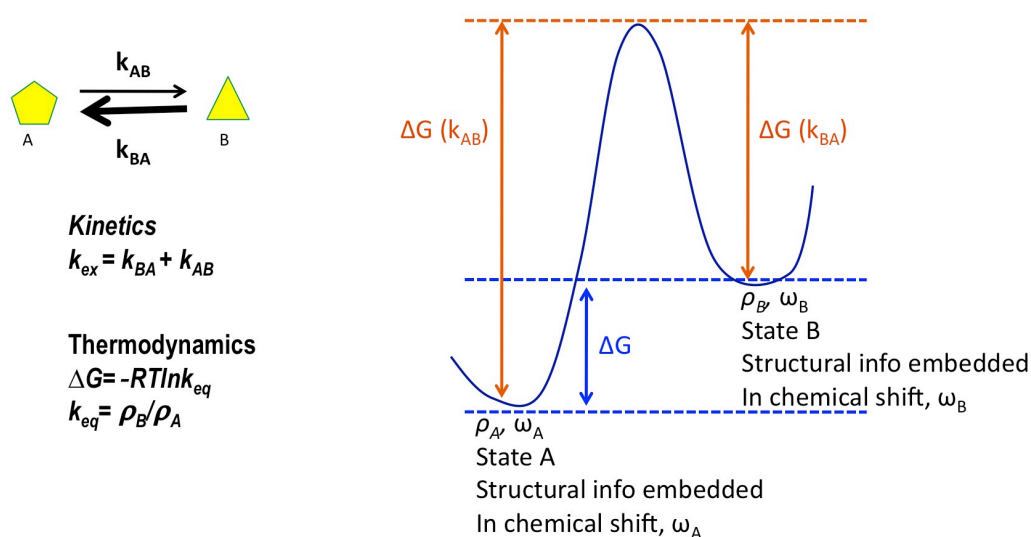


Figure 1.18. Slow timescale NMR dynamics experiments allow measurement of a two-state transition process The slower timescale dynamics (μs - ms), requires chemical shift difference, and are particularly useful to characterize a two-states process A and B (e.g. open-closed conformation. These NMR techniques can measure exchange rates (kinetics) by the relation $k_{\text{ex}} = k_{\text{BA}} + k_{\text{AB}}$, populations (thermodynamics) by measure of the $\Delta G = -RT \ln K_{\text{eq}}$.

Exchange processes may result from conformational flexibility, chemical reactions, and formation of intermolecular complexes. It has important consequences for both the appearance of the spectra and consequently the information that can be derived from it.

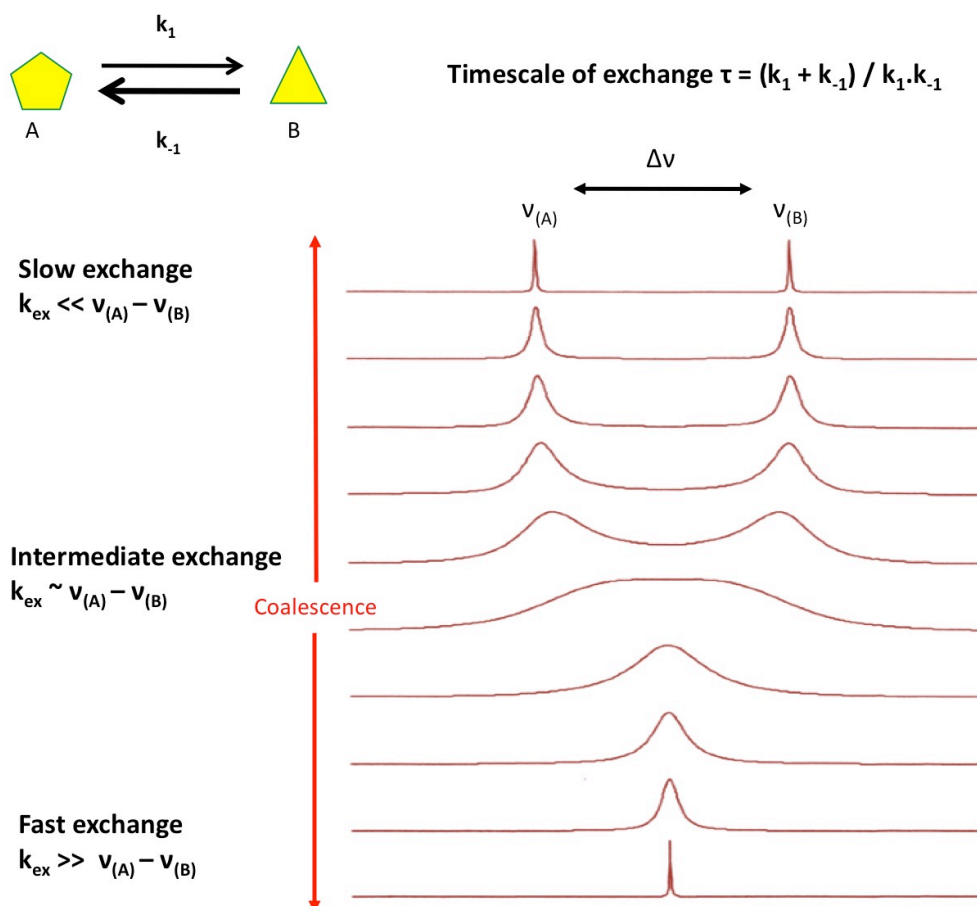


Figure 1.19. NMR spectra for a pair of nuclei exchanging between two states with equal population. For the simplest situation where A and B are inter-converting forms of a molecule, the timescale of exchange (τ) can be measured using the equation illustrated. $\Delta\nu$ represents the difference in frequency (Hz) between the two peaks. The exchange is fast when the rate of interconversion is slow (k_1 and k_{-1} are small compared to $\Delta\nu$), inversely it is fast when the rate of interconversion is fast. In between where the two limits overlap is considered an area of intermediate exchange.

NMR spectral reports of exchange between two intermolecular complexes are a function of the timescale. This is characterised by the $\Delta\nu$ (representing the difference in frequency (Hz) between the two states $\nu_{(A)}$ and $\nu_{(B)}$) and the timescale of exchange as illustrated in fig 1.19. Slow exchange is characterised by $k_{ex} \ll \nu_{(A)} - \nu_{(B)}$ (where k_1 and k_{-1} are small compared to the frequency difference) and the presence of two sharp peaks on the NMR spectra, representative of the two states. Fast exchange is characterised by $k_{ex} \gg \nu_{(A)} - \nu_{(B)}$ as well as the presence of a single sharp peak equidistant from the two signals observed for a slow exchange defined as follow:

$$\delta_{\text{obs}} = \alpha\delta_A + (1 - \alpha) \delta_B \quad \text{Equation 1.14.}$$

Where the δ_{obs} is the chemical shift observed for fast exchange, δ_A and δ_B are the chemical shifts of the two species inter-converting observed at slow exchange, and α is the fractional population of species A for the equilibrium.

In between where the two limits overlap is an area of intermediate exchange characterised by line broadening where $k_{\text{ex}} \sim \nu_{(A)} - \nu_{(B)}$.

In the absence of any coupling and at coalescence conditions (Fig.1.19.), the first-order inter-conversion rate constant can be calculated as followed:

$$k_{\text{TC}} = \pi (\Delta\nu/\sqrt{2}) \quad \text{Equation 1.15.}$$

Where k_{TC} is the rate constant (in s^{-1}), T_c is the coalescence temperature (in Kelvin), and $\Delta\nu$ is the difference in chemical shift between (in Hz) of the two signals without exchange.

This thesis puts forward a strategy for the use of Nuclear Magnetic Resonance spectroscopy together with complementary biophysical methods, to unveil processes of folding, misfolding and polymerisation of α_1 -antitrypsin in unprecedented detail. These processes underlie function and dysfunction in the serpin superfamily of proteins.

Chapter 2 – MATERIALS AND METHODS

2.1 Materials

2.1.1 Regents

All reagents were obtained from Sigma Chemical Company (St. Louis, USA), New England Biolabs (Ipswich, MA, USA), Fisher Scientific UK Ltd or Stratagene (California, USA), unless otherwise stated. BL-21-Gold (DE3) *Escherichia coli* (*E. coli*) were obtained from Stratagene (California, USA). Primers for the mutagenesis experiments were synthesized by Operon (Huntsville, USA). Luria broth was made up from LB broth granules from Merck KGaA (Germany). Proteinase inhibitor tablets were obtained from Hoffmann-La Roche Ltd (Burgess Hill, UK). Snakeskin pleated dialysis membrane 10 kDa molecular weight cut-off was obtained from Perbio Science UK Ltd. (Northumberland, UK), and 5 kDa molecular weight cut-off Vivaspinn filter units were from Viva Science (Stonehouse, UK). SDS-PAGE pre-casted gels were provided by Invitrogen (Paisley, UK). Benchmark protein molecular weight markers were obtained from Invitrogen (Paisley, UK). Plasmid purification was carried using a Qiagen Miniprep kit (Manchester, USA). Polyacrylamide gels were stained using InstantBlue Ultra fast protein stain (Paisley, UK) or Novex InstantBlue Ultra fast protein stain (Cambridge, UK). D₂O and DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) were supplied from Sigma chemical company (St. Louis, USA). For mass spectrometry experiments, buffer exchange was performed using mini gel filtration columns (California, USA).

2.1.2 Instruments

2.1.2.1. Cell purification and characterization

The sonicator was supplied by Sonic Vibracell (Newtown, USA). Cell lysis was performed using the Thermo Electron French Cell Press instrument model FA078, from Thermo Fisher Scientific Inc. (Asheville, USA). Chromatography columns and Akta Prime system were obtained from GE Healthcare Biosciences AB (Uppsala, Sweden). SDS-PAGE running apparatus was obtained from Invitrogen (Paisley, UK). The Native-PAGE running apparatus was obtained from

Bio-Rad Laboratories, Inc. (Hercules, CA, USA). The Transverse Urea Gel (TUG)-PAGE were performed using a Bio-Rad Mini-protein II apparatus, Bio-Rad Laboratories, Inc (Hercules, CA, USA).

2.1.2.2. Circular Dichroism

The 62DS Circular Dichroism (CD) spectrometer was supplied by AVIV (Lakewood, USA). 0.1 mm pathlength quartz cuvette was supplied from Hellma UK Ltd (GB-Southend-on-Sea, UK). Urea titration data were recorded on a JASCO J-810 Circular Dichroism (CD) spectropolarimeter, Jasco Inc. (Easton, MD, USA); Cambridge Institute for Medical Research facility (CIMR).

2.1.2.3. Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectra for α_1 -antitrypsin were recorded using spectrometers operating at ^1H frequencies of 700MHz (Bruker Avance III equipped with 4 channels and deuterium decoupling and a 5 mm $^1\text{H},^{13}\text{C},^{15}\text{N}$ cryo-probe with cooled ^1H and ^{13}C pre-amplifiers; UCL Biomolecular NMR Centre and MRC biomedical NMR Centre, UK); 600MHz (Varian Inova equipped with 4 channels and a 5 mm $^1\text{H},^{13}\text{C},^{15}\text{N}$ coldprobe with a cooled ^1H pre-amplifier; UCL Biomolecular NMR Centre, UK); and 900 MHz (Bruker Avance III equipped with 4 channels and deuterium decoupling and a 5 mm $^1\text{H},^{13}\text{C},^{15}\text{N}$ cryo-probe with cooled ^1H and ^{13}C pre-amplifiers; Lille Pasteur Institute NMR centre, France). To minimize the sample requirements, Shigemi tubes were used (Optimized Shigemi Tube for D_2O 5mm) provided by Bruker (Coventry, UK). Data were processed using NMRpipe¹⁴⁴, CCPN¹⁴⁵ and Sparky software¹⁴⁶.

2.1.2.4. Mass Spectrometry (MS)

Nanomate Robot Triversa for automated sample loading was obtained from Advion (New York, USA). Ions were generated by nanoflow electrospray ionization and spectra acquired on an LCT Premier XE mass Spectrometer

Synapt HDMS by waters UK Ltd (Manchester, UK), quadrupole-TWIMS-oe-ToF Mass Spectrometer, Wolfram Inc. (Oxfordshire, UK). The Mathematica scripts used in the analysis are available at:

<http://people.cryst.bbk.ac.uk/~ubcg66a/software/MSDimer>.

Mass Spectrometry (MS) data were processed using Masslynx 4.1 software Waters Corporation (Milford, USA).

2.2 Methods

2.2.1 Protein recombinant expression in *Escherichia coli* (*E. coli*) and purification

The production of recombinant proteins is a critical avenue for the many structural biology research that require working with relatively large quantities of purified proteins. *Escherichia coli* represents the most commonly used prokaryotic organism for protein expression. It represents a powerful tool for affordable, rapid and abundant protein expression, for which genetics is well established and can be manipulated exhaustively. α_1 -antitrypsin possesses very few basic post-translational modifications, which makes them typically amenable to prokaryotic expression. It is glycosylated within the plasma, however its glycosylation has not been found to be critical to its function or structure⁸⁷. The single cysteine residue in position 232 was found to form a disulfide bond with free cysteine¹⁴⁷ but isn't involved in disulfide bonding within the protein. Therefore, producing recombinant proteins using *E. coli*, is a great tool for structural biology investigation, and furthermore, allows the incorporation of specifically selected isotope labels required for designed solution state NMR experiments.

2.2.1.1. Plasmid constructs

The pTermtat expression vector system (pET) is the most powerful system yet developed for the cloning and expressions of recombinant proteins in *E. coli*. Target genes are cloned in pET plasmids under control of strong bacteriophage

T7 transcription and translation signals; the expression is induced by providing a source of T7 RNA polymerase in the host cell. The T7 RNA polymerase is so selective and active that once induced, almost all of the cells resources are dedicated to gene expression. For mutagenesis and recombinant protein expression purposes, a soluble pQE expression system: pQE-31 plasmid was used as the vector encoding hexahistidine-tagged recombinant wild-type α_1 -antitrypsin cDNA. This plasmid contains both an antibiotic resistance to ampicillin (for selection of transfected cells) and a T7 promoter. The latter enables induction of protein expression by Isopropyl β -D-1-thiogalactopyranoside (IPTG). The α_1 -antitrypsin cDNA constructs also contained an N-terminal hexa-histidine tag, to aid purification of recombinant protein using a Nickel (Ni)-sepharose chromatography column.

For Polymerase chain reaction (PCR) amplification purposes the following coding and reverse primers were used made to a stock solution of 10 μ M;

Forward primer:

ATCTTCTTTAAAGGCAAATGAGAGAGACCCTTTGAAGTCAAGGACACC

Reverse primer:

GGTGTCTTGACTTCAAAAGGGTCTCTCATTTGCCTTTAAAGAAGAT

PCR mixture were set up as follows 1.25 μ l each of forward and reverse primer stocks, 10x pfu buffer (2 mM), dNTP mix (0.5 mM), Pfu turbo (0.05 U/ μ l), plasmid template pQE-31 (2 ng/ μ l), dH₂O to a final volume 50 μ l.

The following PCR conditions were used: 95 °C for 30 s, then 16 cycles of melting (95 °C for 30 s) annealing (55 °C for 1 min) and elongation (68 °C for 10 min). The bacterially derived methylated template DNA was then digested using 1 μ l Dpn-I (0.39 U/ μ l), and Hind-III and run onto agarose gel (1x) for size determination. Integrity of the α_1 -antitrypsin cDNA was confirmed by sequencing.

2.2.1.2 Transformation of BL-21-Gold *E. coli* with plasmid pQE-31 containing α_1 -antitrypsin cDNA

BL-21-Gold (DE3) are the typical system used for over-expression of serpins under a T7 promoter and offer the advantage of bypassing recombinant expression issues due to toxicity, codon bias, and protease sensitivity.

100 μ L BL-21-Gold (DE3) *E. coli*, pretreated by swirling on ice for 10 minutes with 2 μ L of β -MercaptoEthanol (β ME) were transformed with 2 μ L of α_1 -antitrypsin cDNA contained within pQE-31 plasmid. The mixture was then left on ice for 30 minutes prior to delivering a heat shock at 42 °C for 30 s. Cells were then resuscitated for 2 minutes added 500 μ L of LB media prewarmed to 30 °C. The sample was then incubated at 37 °C/225 rpm for 1 hour, and 100 μ L was plated onto an LB/ampicillin plate using aseptic technique. The remainder was spun at 6000 rpm. All the supernatant was discarded apart from the final ~100 μ L, into which the pellet was resuspended before plating onto a second LB/ampicillin plate. Since the pQE-31 plasmid contains an ampicillin-resistance gene, the plates were then incubated at 37 °C overnight, allowing colonies of successfully transformed cells to grow.

2.2.1.3 Unlabelled recombinant α_1 -antitrypsin protein production

A colony of BL-21-Gold *E. coli* containing plasmid coding for α_1 -antitrypsin was picked from an agar/ampicillin plate and grown in 50 ml starter culture of 2xTY (15 g casein peptone, 10 g yeast extract and 5 g sodium chloride, per liter) and 100 μ g/ml ampicillin at 37 °C / 225 rpm overnight. 10 ml of this culture was inoculated into 750 ml of 2xTY and 100 μ g/ml ampicillin at 37 °C / 225 rpm for 3 hours. The cells growth was measured by photometric Optical Density (OD) at 600 nm wavelength (OD₆₀₀). When the OD₆₀₀ reached the range 0.6-0.7 nm, the protein expression was induced by adding IPTG to a final concentration of 1 mM. The incubation was continued for another 4 hours at 30

°C. Then the cells were pelleted at 6,000 rpm for 15 minutes. The pellet was snap-frozen in liquid nitrogen and stored at -80 °C until required for protein purification.

2.2.1.4 Uniform incorporation of ^{15}N - and ^{13}C - ^{15}N - ^2H - isotopic labeling for NMR

For the production of uniformly isotopically labeled α_1 -antitrypsin, the protein was produced in conditions where the only source of nitrogen +/- carbon contained them exclusively in those isotopic forms. This is achieved by growing cells in 1x M9 minimal media (12.8 g/l of Na_2HPO_4 , 3 g/l of KH_2PO_4 , 0.5 g of NaCl), 1 xM9 salts (1 mM MgSO_4 , 50 μM CaCl_2), 1x Basal Medium Eagle Vitamin Solution (Gibco), $^{14}\text{NH}_4\text{Cl}$ Ammonium chloride (1 g/l), 0.4% (w/v) D-glucose (2 g/l) in glass distilled, autoclaved H_2O , and all buffers were 0.2 μ filter sterilized. For ^{15}N -labeling, $^{15}\text{NH}_4\text{Cl}$ was used; while for ^{15}N - ^{13}C - ^2H - labeling, ^{13}C -D₇-Glucose was used as well, and all growth buffers were made up in 100 % D_2O . A colony of BL-21-Gold *E. coli*, containing plasmid coding for α_1 -antitrypsin, was picked from an LB agar/ampicillin plate, and grown in 50 ml of 2xTY and 100 $\mu\text{g}/\text{ml}$ ampicillin, at 37 °C/225 rpm overnight. 4 ml of this culture was inoculated into 200 ml of unlabelled M9 Minimal Media to allow adaptation of *E.coli* to isotopes incorporation. After 6 hours growth, each 15 ml of overnight growth, was inoculated into 2 litres flasks containing 500 ml of M9 Minimal Media. Flasks were incubated at 37 °C/225 rpm. The cells growth was measured at OD₆₀₀ and protein expression was induced and harvested as described in 2.2.1.3.

2.2.2 Purification of α_1 -antitrypsin

The harvested cell pellet was lysed using a cell disruptor. To prevent protein degradation, one pellet of proteinase inhibitor was added per 50 ml of cell harvest. The crude lysate was then ultracentrifuged at 18 k rpm, during 30 min, at 4 °C; and the supernatant was loaded onto a 2 ml high-affinity nickel-sepharose column that had been recently charged before washing to baseline

U.V. absorbance of outflow with Buffer A (all buffers used for purification are being described in Table 2.1).

Nickel-sepharose purification	
Buffer A	20 mM Sodium phosphate, pH 8.0, 20 mM imidazole, 0.05 M NaCl
Buffer B	20 mM Sodium phosphate, pH 8.0, 200 mM imidazole, 0.05 M NaCl
Washing buffer	20 mM Sodium phosphate, pH 8.0, 500 mM imidazole, 0.05 M NaCl
Charging buffer	0.1 M NiSO ₄
Nickel-sepharose purification	
Buffer A	10 mM Tris, pH 7.4, 5 mM EDTA, 1 mM β ME
Buffer B	10 mM Tris, pH 7.4, 0.5 M NaCl, 5 mM EDTA, 1 mM β ME
Washing buffer	10 mM Tris, pH 7.4, 2.5 M NaCl, 5 mM EDTA, 1 mM β ME
Storage	
Protein Storage buffer for NMR	25 mM Sodium phosphate, pH 8.0, 50 mM NaCl, 1 mM EDTA

Table 2.1: Buffers used in protein purification (all 0.22 μ m filter sterilized)

The bound protein was then eluted using a 20-200 mM (100 % Buffer A to 100 % Buffer B) imidazole gradient. The eluted fractions were monitored for the presence of protein by U.V. absorbance spectrometry at OD₂₈₀, and for the presence of protein with molecular weight consistent for α_1 -antitrypsin by 4-12 % Sodium Dodecyl Sulphate (SDS) Polyacrylamide Gel Electrophoresis (PAGE). α_1 -Antitrypsin positive fractions were loaded onto a 5 ml high affinity Q-sepharose column, washing to baseline and eluting using a 0-0.5 M (100 % Buffer A to 100 % Buffer B) NaCl gradient. The presence of α_1 -antitrypsin was assessed by 4-12 % (w/v) SDS-PAGE.

Elution fractions were assessed as previously and fractions positive for intact α_1 -antitrypsin in the absence of co-eluting species were concentrated to the required stock concentration using a prewashed 5 kDa molecular weight cut-off Vivaspın filter membrane. The protein concentration was calculated using measurement of OD₂₈₀. Aliquots were then snap frozen in liquid nitrogen and stored at -80 °C. The proteins used for HDXNMR were lyophilized in aliquots using a Freeze Dryer CoolSafe, ScanVac 110-4 (Yorkshire, UK).

2.2.3 Biophysical and biochemical characterization

There are various forms of inactive α_1 -antitrypsin, including the latent, cleaved, and polymers. Cleaved α_1 -antitrypsin can be easily identified by Sodium Dodecyl Sulphate-Polyacrylamide Gel (SDS-PAGE), since cleavage liberates the C-terminal fragment of approximately 4 KDa. Full-length inactive material may come from either latency or polymerisation. Since the α_1 -antitrypsin polymers are non-covalent they dissociate into monomers in presence of SDS, therefore their characterisation can be made using non denaturing Native-PAGE gels. The latent forms, although inactive, adopt a subtly different form that can not be characterised by SDS-PAGE or Native-PAGE gels. The most accurate way to assess α_1 -antitrypsin activity is to undertake an activity assay to assess the capability of α_1 -antitrypsin to inhibit a target protease.

2.2.3.1 Activity Assay

The activity was calculated as previously described ¹⁴⁸. 5 pmol of active bovine α -chymotrypsin was incubated with increasing concentration of α_1 -antitrypsin (initially estimated active site concentration of 0.1 μ M) in a total volume of 100 μ l reaction buffer (0.03 M sodium phosphate, 160 mM NaCl, 0.1 % (w/v) PEG 4000, pH 7.4).

0.1 μ M α_1 -antitrypsin	1 μ M α -chymotrypsin (in 1 mM HCl, 2 mM CaCl_2)	Reaction Buffer (as described above)
0 μ l	5 μ l	95 μ l
10 μ l	5 μ l	85 μ l
20 μ l	5 μ l	75 μ l
30 μ l	5 μ l	65 μ l
40 μ l	5 μ l	55 μ l
50 μ l	5 μ l	45 μ l

Table 2.2: Buffers used in activity assay (all 0.22 μ m filter sterilized)

The reaction proceeded for $> 5 \times t_{1/2}$ (i.e. > 10 min) at room temperature and the residual proteolytic activity of chymotrypsin was determined by addition of the substrate succinyl-L-alanyl-L-alanyl-propanyl-L-phenylalanyl-p-

nitroanilide to a final concentration of 0.1 mM kept in DMSO. The change of OD₄₀₅ over 3 minutes was observed. Accurate active site concentrations were obtained by plotting residual proteolytic activity against the volume of α_1 -antitrypsin. Since the inhibitory serpin:protease complex has a 1:1 stoichiometry the quantity of active α_1 -antitrypsin was deemed to equal 5 pmol where residual proteolytic activity reached zero.

2.2.3.2 Far U.V. Circular Dichroism spectroscopy

For circular dichroism (CD) spectroscopy, samples were diluted to 3.8 mg/ml (constant buffer conditions: 10 mM Tris, 10 mM NaCl, 1 mM EDTA, pH 7.4, 0.2 μ m filter sterilized) and 50 μ l assessed in a 0.1 mm or 100 μ l assessed in a 0.2 mm pathlength quartz cuvette. Changes in the secondary structure of α_1 -antitrypsin with time and temperature were measured by monitoring the CD signal at 222 nm with the protein at 0.5 mg/ml in 50 mM phosphate, pH 7.4. Thermal unfolding experiments were performed using a heating rate of 60 °C/h.

The urea titration behaviour of α_1 -antitrypsin was studied by CD spectroscopy at a protein concentration low enough to minimize polymerisation effects (0.1 mg/ml, 2 μ M in PBS pH 7.5) in a 0.1 mm path length quartz cuvette at 20 °C. Spectra were recorded between 190 and 260 nm for baseline and data are the mean of 3 repeats of 10 spectral accumulation. Data were cropped from the point where the HT readings exceeded twice the starting value.

2.2.3.3 Native-PAGE

7.5% (v/v) non-denaturing PAGE was used to assess the conformational homogeneity of the α_1 -antitrypsin samples. Separation gel buffer (1.5 M Tris-Cl, pH 8.9) and stacking gel buffer (0.47 M Tris-Cl pH 6.9) were used to prepare the polyacrylamide gel as follow:

	Separation Gel	Stacking Gel
30 % (v/v) Bis/Acrylamide (monomer)	3.75 ml	2.7 ml
Separation Gel Buffer	3.75 ml	-
Stacking Gel Buffer	-	1.6 ml
ddH ₂ O	7.3 ml	10.6 ml
TEMED	30 µl	25 µl
10 % (w/v) Ammonium persulphate (AmPS - added immediately before casting the gel)	60 µl per mix	20 µl per 2 ml

Table 2.3: Buffers used for native-PAGE

Once the AmPS was added, the separation gel mix was pipetted into the gap between the gel casting plate at about $\frac{3}{4}$ of the total volume. The rest of this space was filled with 70 % ethanol and left on the bench for 20-30 min, to allow complete polymerisation of acrylamide. Once polymerised, the ethanol was removed, and the stacking gel solution was poured before inserting a 10 well forming comb. Once this had set samples, diluted in 4x loading buffer, were loaded into each well. They were then discontinuously electrophoresed using cathode buffer (central chamber: 58 mM Tris, 63 mM glycine pH 8.9) and anode buffer (outer chamber: 100 mM Tris, pH 7.8) with constant current (mA) of 90V until the loading dye front was at the bottom of the gel. All proteins were then visualized using InstantBlue.

2.2.3.4 Transverse Urea Gradient-PAGE

A transverse Urea Gradient (TUG) gel is a native polyacrylamide poured gel with a gradient of urea from 0 to 8 M, perpendicular to the direction of electrophoresis¹⁰¹. The 0 M and 8 M urea gel mixes were poured between glass plates that were initially rotated 90 degrees from the orientation used for electrophoresis, using a mixer to generate a gradient from 8 M to 0 M urea. This requires minor modification of most electrophoresis apparatus, including spacers to fit the sides of the square glass plates (as oriented for casting) and bulldog clips for holding the plates in this orientation. 1% agarose solution was used to seal the gel to avoid leakage during the gradient process. 500 µl of 8 M urea gel mix were used before the start of the gradient, and 500 µl of 0M urea gel mix were used to finish. Once polymerised, the gel was rotated 90 degrees back to allow pouring of the stacking solution, and insertion of a single well comb (0.1

mm thick, and 4 cm wide well). Samples were diluted in 4x loading dye, and 40-50 µg were loaded into the stacking gel and discontinuously electrophoresed using cathode (upper reservoir) buffer (58 mM Tris, 63 mM glycine, pH 8.9) and anode (lower reservoir) buffer (bottom of the gel: 100 mM Tris, pH 7.8) with constant current (25 mA per gel).

	0 M urea	8 M urea
Non-denaturing PAGE separation buffer	12 ml	12 ml
Acrylamide	11.2 ml	11.2 ml
Urea	-	23 g
Water	To final volume of 48 ml	To final volume of 48 ml
TEMED	50 µl	50 µl
10 % (w/v) AmPS	20 µl per 3.45 ml	20 µl per 3.45 ml

Table 2.4: Buffers used for TUG-PAGE

2.2.3.5 ¹H NMR

¹H 1D experiments were run with excitation sculpting for water suppression, with an acquisition time of 50 ms and a spectral width of 11 ppm. The number of scans varied from 128 to 512 typically with a recycling delay of 1s.

2.2.3.6 Native Mass Spectrometry (MS)

The TWIMS-MS experiments were carried out on a Synapt High Resolution Mass Spectrometry (HRMS) quadrupole-Traveling Wave Ion Mobility Mass Spectrometry - orthogonal acceleration – Time of Flight (TWIMS-oa-ToF). It was calibrated for mass:charge (m/z) ratios, using 33 µM cesium iodide in 250 mM ammonium acetate. 3 µl samples were loaded onto nano-electrospray ionization (ESI) needles made in-house, using a previously described method¹⁴⁹. Conditions were optimized, samples were run under “mobility-time of flight” (MOB-TOF) mode, and spectra were acquired for at least 3 minutes (equivalent to 90 scans). Spectra were calibrated externally using an aqueous 33 µM solution of caesium iodide, in 250 mM ammonium acetate, as previously described by the

Ashcroft (University of Leeds, UK), and Robinson (Oxford University, UK) groups¹⁴⁹. All experiments were undertaken under identical voltage conditions: capillary voltage 0.6 kV; sampling cone voltage 35 V; trap energy 10 V. They were repeated over a range of voltage wave heights, with transfer energies of either 10 or 30 V for non-dissociative (native) MS experiments. For the Ion Mobility - Mass Spectrometry (IM-MS) experiments collision energy, transfer energy and trap energy were all 60 V each with other conditions held constant.

2.2.4 Hydrogen Deuterium exchange (HDXNMR)

2.2.4.1 Sample preparation

α_1 -antitrypsin samples for HD exchange were prepared at 250 μ M concentration (300 μ l), in 100 mM ammonium bicarbonate before being snap frozen and lyophilized overnight. Buffer used to dissolve the protein while recording the NMR experiment (20 mM sodium phosphate, 50 mM NaCl, 1 mM EDTA, d H₂O (0.2 μ m filter sterilized)), was produced in a range of pH values: 8.0, 7.0, 6.5, 6.0, and 5.75; and was also lyophilized using the same protocol. The buffer was then rediluted in 100 % D₂O (containing 0.001 % DSS for spectra reference) to give equivalent pD* (8.4, 7.4, 6.9, 6.4, and 6.15 respectively). Once the NMR buffer was dissolved in 100 % D₂O, it was used to dissolve the sample of lyophilized α_1 -antitrypsin to 250 μ M final concentration. The sample was then immediately used to record the HD exchange experiments. Dead times were recorded (5 to 7 minutes range).

2.2.4.2 ¹H-¹⁵N SOFAST HMQC NMR for HDXNMR experiments

The 2D ¹H-¹⁵N band-Selective Optimized Flip Angle Short Transient (SOFAST) Heteronuclear Multiple Quantum (HMQC) NMR experiments^{138; 150} used were run with a ¹⁵N spectral width of 32 ppm (64 increments) and a ¹H spectral width of 16 ppm. The excitation pulse (the ¹H 90 degree pulse) was calibrated experimentally on each sample, to ensure optimal proton excitation. Acquisitions were typically of ≥ 4 transients with a 4 transient experiment

lasting 4 minutes. Reference spectra in H₂O were acquired under similar experimental conditions (of concentration, preparation, pH, and temperature) over 4 hours SOFAST-HMQC. For HDXNMR experiments, a 250 µM α₁-antitrypsin concentration was used in order to permit collection of satisfactory spectra in 4 minutes in 2D NMR experiments. ¹H-¹⁵N two-dimensional SOFAST-HMQC spectra were processed and analyzed using nmrPipe and CCPN. All spectra were referenced to DSS at 0.0 ppm, manually phased and baseline corrected.

2.2.4.3. Spectral processing and NMR frequency referencing

The data obtained at 600 MHz (~ 150 SOFAST-HMQC experiments) and 700 MHz (5 sets of ~ 250 SOFAST experiments) were converted from Varian and Bruker files, respectively, to NmrPipe readable files. The first spectrum of each set of HD analysis was processed using NmrPipe, by multiplying the FID by a multiplier or window function to reduce the noise. We applied a filter solvent and a baseline correction before Fourier transforming the time domain data to obtain a frequency domain spectrum. The correction of the phase error was done by adjusting the phase of the proton, and of the nitrogen dimensions, where necessary. The desired region (in ppm) of the full spectral window was selected. Each HD exchange dataset included 1D NMR spectra recorded 30 minutes post-dead time, and one per day until the completion of the experiment. 1D spectra were processed using the Bruker command: *Set XCAR*, and then calibrated on the δDSS chemical shift. The calibrated values δSF1 and δSF3 used for referencing the 2D experiments were calculated from the following relations (for which the BF1, SF1, and SF3 values are known and found in the *acquus file*):

$$\begin{aligned}\nu_{0,H} &= BF1 \times \left(1 + \frac{\delta_{DSS}}{10^6}\right) \\ \nu_{0,N} &= 0.101 \times \nu_{0,H} \\ \delta_{SF1} &= \frac{SF1 - \nu_{0,H}}{\nu_{0,H}} \times 10^6 \\ \delta_{SF3} &= \frac{SF3 - \nu_{0,N}}{\nu_{0,N}} \times 10^6\end{aligned}$$

2.2.4.4. Data processing and analysis

A detailed analysis has been conducted of changes in protection from HDXNMR in ≤ 200 residues with discrete, unambiguous cross-peaks in well-dispersed regions of ^1H - ^{15}N spectra across a pH titration using two-dimensional real-time SOFAST Heteronuclear Multiple Quantum Correlation ^{138; 150} (HMQC) NMR spectroscopy (pulse program and the typical number of scan patterns used for each pH condition for the HD exchange). Experiments were run with a ^{15}N spectral width of 32 ppm (64 increments) and a ^1H spectral width of 16 ppm. Spectra were processed and analyzed using *nmrPipe* and *CCPN suites*. HDXNMR was measured using a uniformly ^{15}N labeled, uniformly protonated sample, that was lyophilized and re-dissolved in 100 % D_2O containing 0.001 % DSS for spectra reference with no other change to buffer conditions, at pH 8.0, pH 7.0, pH 6.5, pH 6.0, and pH 5.75. The protein concentration was 250 μM and experiments were conducted at various magnetic field strengths (600, 700, 800 MHz spectrometers).

For spectral processing, the referencing and processing of the first spectrum was undertaken manually. A python script (courtesy of Dr Chris Waudby, UCL) was then generated to process, and reference all spectra in a similar manner, including the appropriate phasing and referencing. This generated a list of ^1H - ^{15}N SOFAST HMQC spectra, and combined sequential scans to give a '3D experiment' (with time as 3rd dimension) using the 'series.com' and 'xyz2pipe' *nmrPipe* commands. The resulting spectrum was then opened using the *CCPN suite*¹⁵¹. A crosspeak list was generated from the first scan of each HDXNMR experiment (the list comprised ~ 200 crosspeaks) and resonance assignments were attributed by comparison with the assignment spectra recorded at pH 8.0 and pH 7.0¹⁴⁰, before being applied to the rest of the HDXNMR in each pH conditions.

Backbone amide exchange is reported as exponential decline in intensity according to Equation 2.1:

$$I(t) = I_0 \cdot e^{(-k_{ex}t)} + c \quad \text{Equation 2.1}$$

where $I(t)$ is the intensity at time t , I_0 is the intensity at time 0, and k_{ex} is the rate of exchange.

k_{ex} values were therefore calculated by fitting curves of equations of the form $y = a \cdot e^{(-b \cdot x)} + c$ to data for residues that demonstrated intensity decline, using the program *cftool* in the *Matlab suite*⁸⁸. HDXNMR is related to H/D ion availability and therefore the pH value and the measured pD* values differ by 0.4 units¹⁵².

$$pH = pD^* - 0.4 \quad \text{Equation 2.2}$$

HD exchange is pH independent at high values of pH (EX1 regime) but entirely pH-dependent (i.e. direct inverse-relationship to H⁺/D⁺ ion availability) at low pH values (EX2 regime). Under EX2 conditions k_{ex} can be readily converted to a protection factor (P) using Equation 2.3:

$$P = k_{int}/k_{ex} \quad \text{Equation 2.3}$$

Where k_{int} is the experimentally-derived, unprotected backbone amide hydrogen exchange rate (as stored for all residues within SPHERE program¹⁵³).

An EX2 regime is observed at pH values up to and including a range where variation by 1.0 pH unit (10-fold change in deuterium ion availability) changes the k_{ex} value 10-fold in the same direction as the pH. For a pH change of 0.5 in EX2 the change is $\sqrt{10}$ -fold i.e. 3.16. Similarly, in the case of 0.25 and 0.75 pH unit in EX2, the k_{ex} ratios calculations the change should be equal to 5.62, and 1.77, respectively.

Previous studies have indicated that α_1 -antitrypsin begins to undergo acid or base-induced polymerisation between pH 6.0 and 5.0, and between pH 8.0 and 9.0, respectively. The titration was therefore performed across the range pH 6.0 - 8.0 (equivalent to pD* 6.4 - pD* 8.4). The pH 5.75 (pD* 6.25) was also monitored, coupled with off-pathway polymerisation characterisation. pH values were determined using a pH-meter with NMR tube probe.

Exponential fitting of the intensity decays over time provides information on the relative flexibilities of ≤ 200 residues within the native molecule in solution. The use of *Matlab* software for fits of peaks intensity decays.

Finally, P relates to the difference in free energy of the observed residue, between the “closed” state and the “open”, exchange-competent state ΔG_{op-cl} ¹⁵⁴.

$$\Delta G_{op-cl} = RT \ln(P) \text{ or } \Delta G_{app} = -RT \ln(k_{op}) \quad \text{Equation 2.4}$$

where R is $1.985877534 \cdot 10^{-3} \text{ kcal.K}^{-1}.\text{mol}^{-1}$ and T is 298K

2.2.5. NMR studies of α_1 -antitrypsin in urea

2.2.5.1. Sample preparation

NMR spectroscopy of α_1 -antitrypsin across a urea titration was performed at $100 \mu\text{M}$ protein concentration. This permitted collection of satisfactory spectra in 1-4 h for 2D. Spectra were recorded in the buffer described above, to which 10 % D_2O was added, as a spectrometer field lock, and 0.001 % DSS was added for referencing.

Samples were prepared from $250 \mu\text{M}$ α_1 -antitrypsin stock, and different volumes of a 9.0 M urea stock to give the appropriate final urea concentrations 7.6 M, 6.0 M, 4.0 M, 3.0 M, 2.0 M, 1.5 M, 1.0 M and 0.0 M. Urea concentrations were verified by refractometry. The homogeneity of the sample was confirmed by SDS-, non-denaturing- and transverse urea gradient (TUG)-PAGE and activity assay. The conformational state of the sample following NMR experiments was assessed by far U.V. circular dichroism spectroscopy.

2.2.5.2 ^1H - ^{15}N -TROSY HSQC

Transverse Relaxation Optimized Spectroscopy (TROSY) - Heteronuclear Single Quantum (HSQC) NMR experiments¹³¹ were used for the urea equilibrium studies. This experiment is useful for study of slow tumbling species enabling significant sensitivity improvements. The pulse sequence used allowed water suppression by suppression of radiation damping. ^1H - ^{15}N -TROSY-HSQC experiments were recorded using enhanced gradient selection to minimise breakthrough from unlabelled urea. The spectra were referenced and processed as described in the section 2.2.4.3, and 2.2.4.4.

2.2.6. NMR relaxation of α_1 -antitrypsin

2.2.6.1 Sample preparation

Perdeuterated, uniformly ^{15}N , ^{13}C -labelled α_1 -antitrypsin sample was prepared as described in section 2.2.1.4. In order to allow all the backbone nuclei to back exchange with the protonated solvent for observation in relaxation experiments, the sample was subjected to an unfolding/refolding cycle as follows: the protein sample was concentrated down to 500 μl (300 μM concentration) into unfolding buffer (8.0 M Urea, 50 mM Tris, pH 8.0, 100 mM dithiothreitol (DTT)). The unfolded protein were then refolded into 500 ml refolding buffer (25 mM Sodium phosphate pH 7.0, 50 mM NaCl, 1 mM EDTA) drop by drop with continuous mixing on magnetic stirrer at 4 $^\circ\text{C}$, overnight, before being concentrated as described in 2.2.1.3.

The resulting sample was concentrated to 250 μM and was characterized biochemically via SDS-PAGE, native-PAGE, and activity assay, as well as 1D ^1H -NMR spectroscopy, before being used for NMR relaxation experiments.

Double labelled sample was prepared as described in section 2.2.1.4 in a fully deuterated environment. In order to allow all the backbone nuclei to back exchange with the protonated solvent for relaxation experiments, the sample

was unfolded and refolded. The protein sample was concentrated down to 500 μ l (300 μ M concentration) into unfolding buffer (8.0 M Urea, 50 mM Tris, pH 8.0, 100 mM dithiothreitol (DTT)). The unfolded protein were then refolded into 500 ml refolding buffer (25 mM Sodium phosphate pH 7.0, 50 mM NaCl, 1 mM EDTA) drop by drop with continuous mixing on magnetic stirrer at 4 $^{\circ}$ C, overnight, before being concentrated as described in 2.2.1.3. The sample was concentrated to 250 μ M, and characterized biochemically by SDS-PAGE, native-PAGE, and activity assay, as well as 1D 1 H- NMR spectroscopy, before being used for NMR relaxation experiments.

**Chapter 3 – NMR STUDIES OF α_1 -ANTITRYPSIN SOLUTION
BEHAVIOUR IN UREA**

3.1. Introduction

3.1.1. Metastability/Z-mutant and formation of intermediate ensemble.

The folding, misfolding, and inhibitory mechanisms of serpins are linked to both thermodynamic metastability and conformational flexibility. Metastability of native α_1 -antitrypsin arises from inefficient structural packing around the fold. It is most strikingly associated with a reactive loop motif that is minimally-stabilized in solution and a parallel β -strand packing between s3A and s5A within an otherwise antiparallel sheet. Such structural features predispose to conformational change resulting in a thermodynamic hyperstabilisation to form a stable serpin-protease complex during inhibition or equivalent S to R transitions¹⁵⁵. In general, the metastable native state is also believed to be more dynamic (kinetically unstable) in solution than hyperstabilised states⁹².

Homozygosity for the Z allele of α_1 -antitrypsin accounts for almost 95% of all cases of α_1 -antitrypsin deficiency⁸⁸. The Z (Glu342Lys) mutation occurs at the head of strand 5 of β -sheet A and the base of the RCL. It has been variously suggested that it stabilizes a folding and/or an intermediate S to R transition state of α_1 -antitrypsin, and that it reduces the thermodynamic and/or kinetic stability of its native fold^{103; 156}. It is believed to slow the folding of the α_1 -antitrypsin molecule^{157; 148}, and so increase the population of a conformer that allows the RCL of one molecule to insert into β -sheet A in the absence of proteolytic cleavage, giving rise to an intermolecular linkage that can be propagated within chains of polymers^{95; 158}. The characterization of the structural basis of α_1 -antitrypsin deficiency therefore requires understanding of the folding pathway, how this relates to formation of polymeric and latent conformers, and the mechanisms by which these species are handled *in vivo*. However, detailed characterization of stability and flexibility in serpins in solution using structural methods is challenging due to their relatively large size and the concomitant propensity for polymerisation/aggregation.

Many structural biology experiments are conducted using recombinant non-glycosylated proteins. However α_1 -antitrypsin is glycosylated *in vivo*, so it is useful to consider how glycosylation affects the protein folding and function, in order to determine whether findings are applicable to *in vivo* systems. Indeed, glycosylation usually increases the global stability of proteins, and further, global stabilization is often accompanied by reduced flexibility in the native state¹⁵⁹. However, no apparent discrepancy has been observed between the kinetic and thermodynamic stability in wild type (M) α_1 -antitrypsin in the glycosylated form observed *in vivo*, and its recombinant counterpart¹⁰⁰. This indicates that further structural exploration of the recombinant non-glycosylated molecule can be relevant to the *in vivo* context.

3.1.2. Equilibrium folding intermediate

Various biochemical and biophysical techniques have been used to separate and characterize proteins with different oligomerization states and/or conformations formed by perturbation of the native state. Examples of such perturbations are those induced by changes in solution conditions, e.g. temperature or pH. The most physio-pathologically relevant perturbations are those induced in the context of mutations found in patients presenting with plasma deficiency¹⁶⁰ under at physiological pH. However monomeric Z α_1 -antitrypsin cannot be purified above trace yield from recombinant *E. coli* systems. Alternatively, denaturant studies (e.g. titrations, kinetic studies) are widely used to probe protein folding and unfolding pathways¹⁶¹. The two most commonly used denaturants are guanidine and urea. The mechanism by which they destabilize a given protein structure is controversial. Although structurally similar, urea and guanidinium denature proteins by different mechanisms: urea, unlike guanidine, promotes denaturation via formation of hydrogen bonds with polypeptides¹⁶².

The polypeptide chains of many denatured proteins spontaneously refold to their native conformation in free solution by adjusting solvent conditions to

near physiological¹⁶³. For relatively small globular proteins, the native to denatured state transition involves no other thermodynamically stabilized states (two-state transition) and is symmetrically reversible¹⁶⁴. Two-state unfolding transitions may be characterized by:

- Cooperative changes in protein conformation signals that follow a sigmoidal curve along a denaturant titration;
- Consistent changes observed by multiple spectroscopic methods;
- Observation of an isosbestic point across a denaturant titration by spectroscopy. If the spectra of a protein that is in entirely non-denaturing and entirely denaturing conditions intersect at a specific wavelength (i.e. the two conformational states have the same molar absorptivity (ϵ)) this is termed an isosbestic point. If the transition is two-state, all spectra recorded at partially-denaturing conditions should also pass through this point;
- The measurement of enthalpy value (ΔH_{cal}), obtained using differential scanning calorimetry (DSC) measurements, calculated from the area of the transition peak, should match with the van't Hoff enthalpy value (ΔH_{vH}), calculated under the assumption of a two-state transition. This can be done by comparing the calorimetric enthalpy of denaturation (the area under the peak) A_{peak} to the van't Hoff enthalpy described as follow:

$$\Delta H_{vH}(T) = -R \frac{d \ln K}{dT^{-1}}$$

At $T = T_d$ (temperature at which half the molecule in the system are unfolded) the $\Delta H_{vH}(T_d)$ can be described as:

$$\Delta H_{vH}(T_d) = \frac{RT_d^2 \Delta C_p^{max}}{A_{peak}}$$

When a two-state unfolding is observed $A_{peak} = \Delta H_{vH}(T_d)$. The ΔC_p^{max} is the height of the heat capacity peak.

However many proteins, particularly those of higher molecular mass, do not follow these patterns. Instead they fold/unfold via intermediate ensembles that are thermodynamically stabilized enough to be stably populated over a mid-range of denaturant conditions.

Denaturant studies have previously been considered particularly relevant in studying the relationship of these processes to disease-associated polymerisation in systems where low concentrations of denaturant can induce the formation of polymeric states^{99; 161}. However for serpinopathies this has been called into question by the observation that denaturant-induced polymers are poorly recognized by the 2C1 monoclonal antibody that very sensitively detects *ex vivo* and heat-induced polymers⁹³.

3.1.3. Probing the folding/unfolding pathway of α_1 -antitrypsin in guanidine

Over 20 years ago, the first study of α_1 -antitrypsin unfolding was published by performing equilibrium studies with increasing concentration of denaturant (guanidinium hydrochloride)¹⁶⁵. It was performed on glycosylated plasma α_1 -antitrypsin using far-ultraviolet circular dichroism, highlighting the presence of an intermediate ensemble using equilibrium unfolding studies. The authors concluded a three state transition was observed during unfolding: native (N), intermediate (I), unfolded (U). The data were consistent with maximal population of an intermediate ensemble around 1.5 M guanidine⁸³. Interestingly, the polymers that form as a consequence of treatment with both urea and guanidine, are not recognized by the 2C1 monoclonal antibody (mAb), that recognizes the pathological polymers that form *in vivo*.

A recent study by Krishnan *et al.*, explored species populated by α_1 -antitrypsin during early guanidine-induced unfolding by studying a series of single cysteine variants. The cysteine residues, engineered at various sites around the molecule, could be tagged by maleimide moieties if they were accessible to solvent. By quantifying the degree of maleimide tagging in low concentrations of GdnHCl for each variant, the authors proposed this showed intermediate state formation to be accompanied by local unfolding of the C-terminal half of the A-sheet, s5A-hI-s6A. They further proposed that, concertedly, the F-helix unwinds and dissociates from its packing interactions with the N-

terminal strands of β -sheet A, and interactions of the RCL with the B- and C-sheets are disrupted. In contrast strand 3 A, helix B and most of β -sheets B and C were stable. These conclusions fitted very well with the β -hairpin model of polymerisation that had been published prior to this, which was itself in part based upon selective cleavage studies of motif destabilization in low concentrations of GdnHCl⁹⁹. Together these findings contribute to a structural concept of guanidine-induced denaturation proceeding via a conformer resembling the putative intermediate in the β -hairpin linkage model. However both the selective cleavage and maleimide labeling strategies are limited in that they will stabilize the states they report and so skew conformational equilibria to increase their population. Moreover they are interpreted on the assumption that all regions reporting destabilization within the ensemble in this way will be destabilized together within a single intermediate state.

Furthermore these findings must also be taken in the context of other recent studies that have shown α_1 -antitrypsin can adopt different types of polymers under different denaturing concentration⁹³. These may arise from subtle conformational differences in transiently populated structures adopted early in the polymerisation pathway¹⁶⁶.

Guanidine cannot be used for NMR studies of protein denaturation because (i) it creates high viscosity giving rise to line broadening, (ii) high dynamic range (higher concentrations of guanidine, relative to that of the protein requires to apply guanidine suppression as well as the usual water suppression used for NMR experiments), lastly, (iii) chemical exchange (the hydrogens attached to the nitrogens of Gu.HCl exchange with the solvent hydrogens, and the rate of exchange is very high. This gives rise to complications for pulse sequences based on selective excitation as well as for the spectral editing pulse sequences).

3.1.4. Probing the folding/unfolding pathway of α_1 -antitrypsin in urea

Equilibrium-unfolding studies of α_1 -antitrypsin in urea have been undertaken in a series of studies by the Bottomley group^{81; 100; 167}. These data supported the presence of a single intermediate around 4.0 M urea⁸¹. Tryptophan mutagenesis and fluorescence studies indicated that, during urea-induced unfolding, dynamic changes occurred around β -sheet A consistent with its expansion in the intermediate state⁸⁸. More limited structural change was observed from a tryptophan residue reporting on the environment around β -sheet B with evidence of some residual secondary structure even at 8.0 M urea. Kinetic studies of wild-type α_1 -antitrypsin in urea have further supported the involvement of 2 kinetic intermediates along the polymerisation pathway⁸⁹.

Transverse urea gradient (TUG) PAGE distinguishes components based on their relative stability to increasing concentrations of urea. The native conformation of plasma α_1 -antitrypsin is compact, and so migrates relatively rapidly at conditions < 1.0 M urea; however, at higher concentrations of urea it undergoes a transition to a denatured, poorly migrating species. Stabilised monomeric states (latent or cleaved) or enzyme complexed species remain stable in their compact conformations across the whole gradient⁸⁸. Heat-induced and *ex vivo* polymers are also hyperstable in urea, retaining their characteristic ladder appearance across the entire gradient⁸⁸. A complex unfolding profile has been observed previously by TUG PAGE for the Z variant compared to the M α_1 -antitrypsin that was taken to demonstrate a folding defect induced by the mutation¹⁶⁸. However this has not been reproduced in other hands (Lomas and Bottomley groups, personal communication), perhaps reflecting variability in techniques used to form the urea gradient.

3.1.5. Effect of stabilizing 'cavity-filling' mutations on the native state α_1 -antitrypsin

A number of studies have been undertaken to investigate the relationship between the inhibitory mechanism and the thermodynamic stability of α_1 -antitrypsin. It has been proposed that the metastability of native serpins arises from suboptimal structural packing that facilitates their transition to hyperstable inhibitory complex in which packing is improved¹⁵⁵. Consequently, mutations aiming to improve side-chain packing are expected to increase the thermodynamic stability of α_1 -antitrypsin and reduce inhibitory function. One such mutation, Gly117Phe - predicted to occupy the cavity between s2A and the F-helix, exemplifies these consequences, and together with other variants validated these concepts¹⁶⁹ when characterised biochemically (Fig.3.1.). The subsequent crystal structure obtained for the Gly117Phe mutant demonstrated no significant conformational change other than a downward shift in the position of the F-helix¹⁷⁰. Dynamic effects of this mutation were further assessed by HDXMS studies. These showed changes in regions involved in inhibitory mechanism particularly in the C-term part of F-helix. Interestingly, however the conformational changes were not confined locally but instead were spread to key regions involved in the inhibitory mechanism such as the RCL and β -sheet A, as well as β -sheet C and helix D. Similarly, mutagenesis to introduce a space-filling phenylalanine residue within the cavity flanking β -sheet A was also tested at the residue Thr114. Indeed, Thr114Phe significantly raised the melting temperature of α_1 -antitrypsin and was shown to delay polymer formation¹⁶⁹. These data demonstrate the importance of this cavity as a potential drug target to ameliorate polymerization.

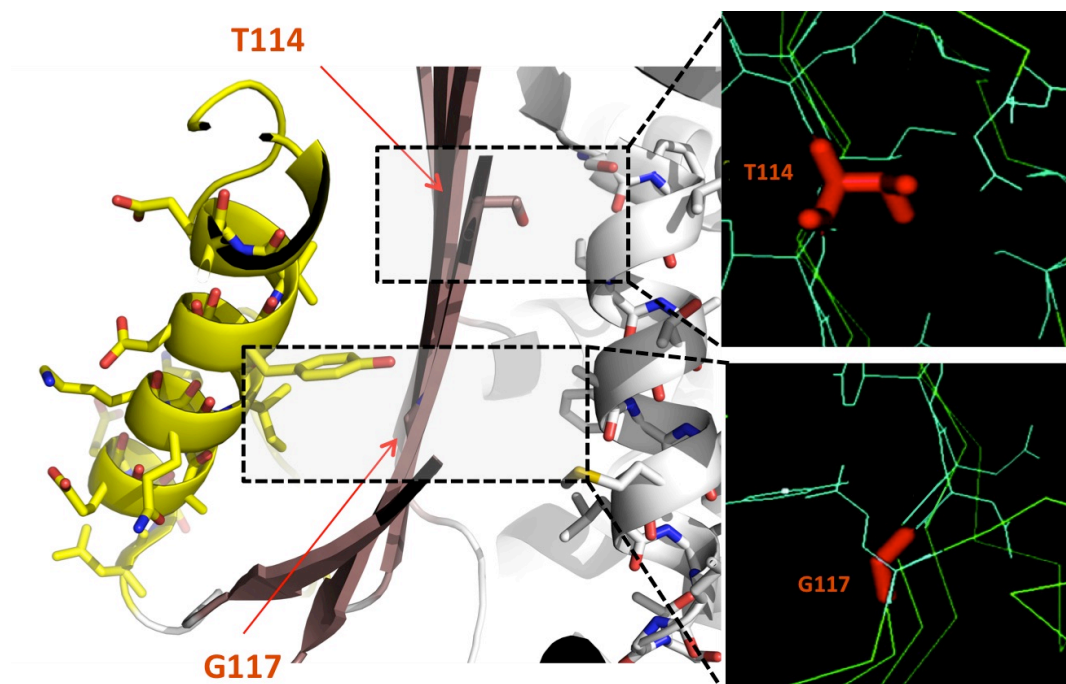


Figure 3.1. Crystal structure of monomeric α_1 -antitrypsin illustrating the environment of the residues Thr-114 and Gly-117. In the left panel, the β -sheet A (brown), the F-helix (yellow) (PDB:NE4). The hydrophobic surface cavity of interest is bounded by strand 2 of β -sheet A (s2A), helix B (hB), helix D (hD) and helix E (hE). This area is obliterated during conformational transitions that involve reactive loop insertion into β -sheet A. The right panel represents a model of the interior of the hydrophobic cavity displaying the position of the amino acid side chains (blue), readapted from¹⁶⁹. The residues Thr-114 and Gly-117 (red) on s2A were highlighted.

I therefore undertook unfolding equilibrium studies using urea titration of the native state α_1 -antitrypsin using NMR and CD spectroscopy to understand the equilibrium behaviour of α_1 -antitrypsin in denaturant and provide more detailed information on the previously described intermediate state(s). This was complemented by additional stability studies using CD spectroscopy of the WT, M, Thr114Phe, Gly117Phe, and Z- α_1 -antitrypsin conformers.

3.2 Results

3.2.1. NMR studies of α_1 -antitrypsin unfolding structural behaviour

The ^1H - ^{15}N TROSY-HSQC NMR spectrum of α_1 -antitrypsin in the absence of urea (0 M) was as described previously¹⁴⁰ and provided a reference for spectra collected at low urea concentration where assignable cross-peaks could be observed (Fig. 3.2).

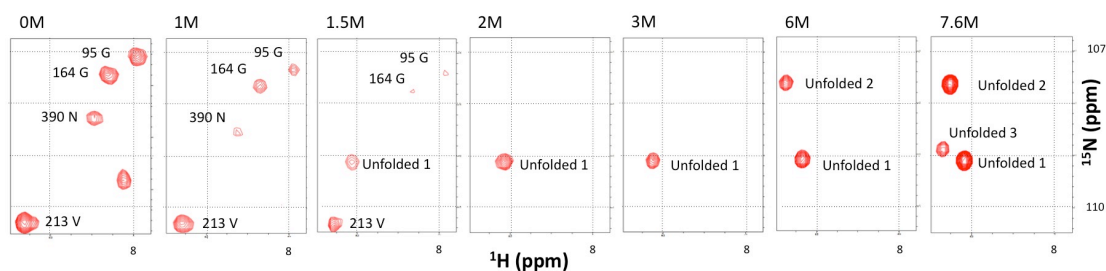


Figure 3.2. ^1H - ^{15}N -TROSY HSQC NMR spectroscopy urea titration for recombinant α_1 -antitrypsin. ^1H - ^{15}N - TROSY HSQC NMR spectroscopy spectra were recorded across the titration between 0.0 M and 7.6 M urea. This illustrate the use of assignment at 0.0 M urea to follow equilibrium unfolding behaviour, as urea concentration increases those peaks loose their intensity while new unfolding peaks appear at higher urea concentration.

Figure 3.3., shows α_1 -antitrypsin fingerprints across the titration, where the 0 M urea spectrum represents a fully-folded monomeric sample with well-dispersed resonances, and the 8 M urea theoretically contains fully-unfolded proteins with resonances appearing in the 8.5–7.5 ppm range in the ^1H proton dimension.

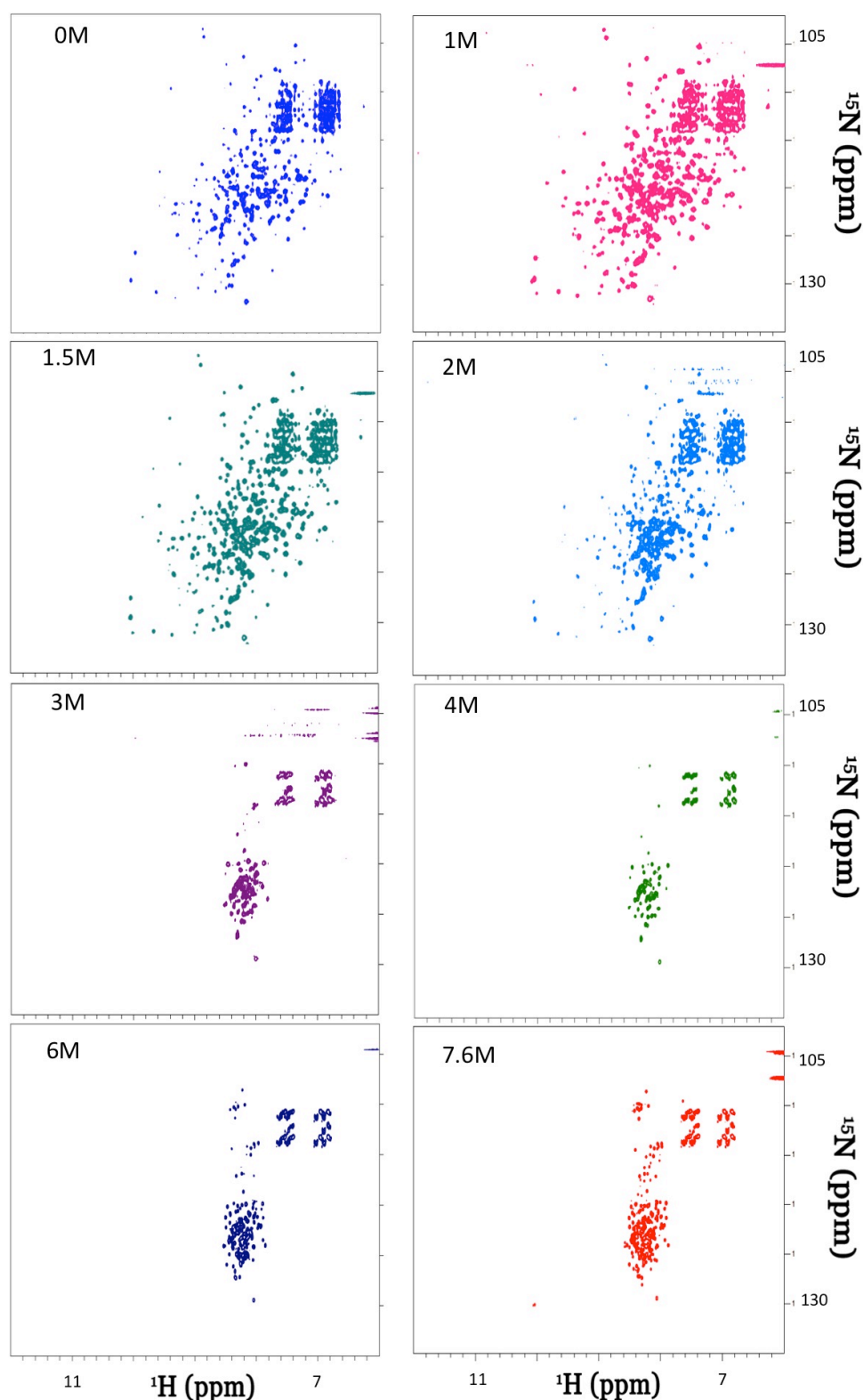


Figure 3.3. ^1H - ^{15}N -TROSY HSQC NMR spectroscopy urea titration for recombinant α_1 -antitrypsin. ^1H - ^{15}N -TROSY HSQC NMR spectroscopy spectra were recorded across the titration at 0M, 1.0M, 1.5M, 2.0M, 3.0M, 4.0M, 6.0M, and 7.6M urea. The 0M urea spectrum represents a fully-folded monomeric sample with well-dispersed resonances, while the 7.6M urea theoretically contains fully-unfolded proteins with resonances appearing in the 8.5–7.5 ppm range in the ^1H proton dimension.

In order to monitor any urea-induced off-pathway polymerisation a series of ^1H NMR experiments have been recorded across the titration as illustrated in figure 3.4. Only the samples at 0 M, 1.0 M, 1.5 M, 2.0 M, and 2.5 M urea could provide a suitable signal, as higher concentration of urea disturbed the NMR signal. In order to quantify urea induced polymerisation, intensities were measured at two regions of the spectra: (i) upfield (9.61-9.56, 9.44-9.36, 9.37-9.29 ppm), and (ii) downfield (0.29-0.39, 0.5-0.58, 0.66-0.77 ppm). At higher urea concentrations 1D spectra supported similarly rapid equilibration to that seen at 0 M and 1.0 M urea. Lower right panel shows relative intensities for these conditions in terms of 1D projections of the 2D spectra onto the ^1H axis (black: 4.0 M; red: 6.0 M; blue: 7.6 M urea).

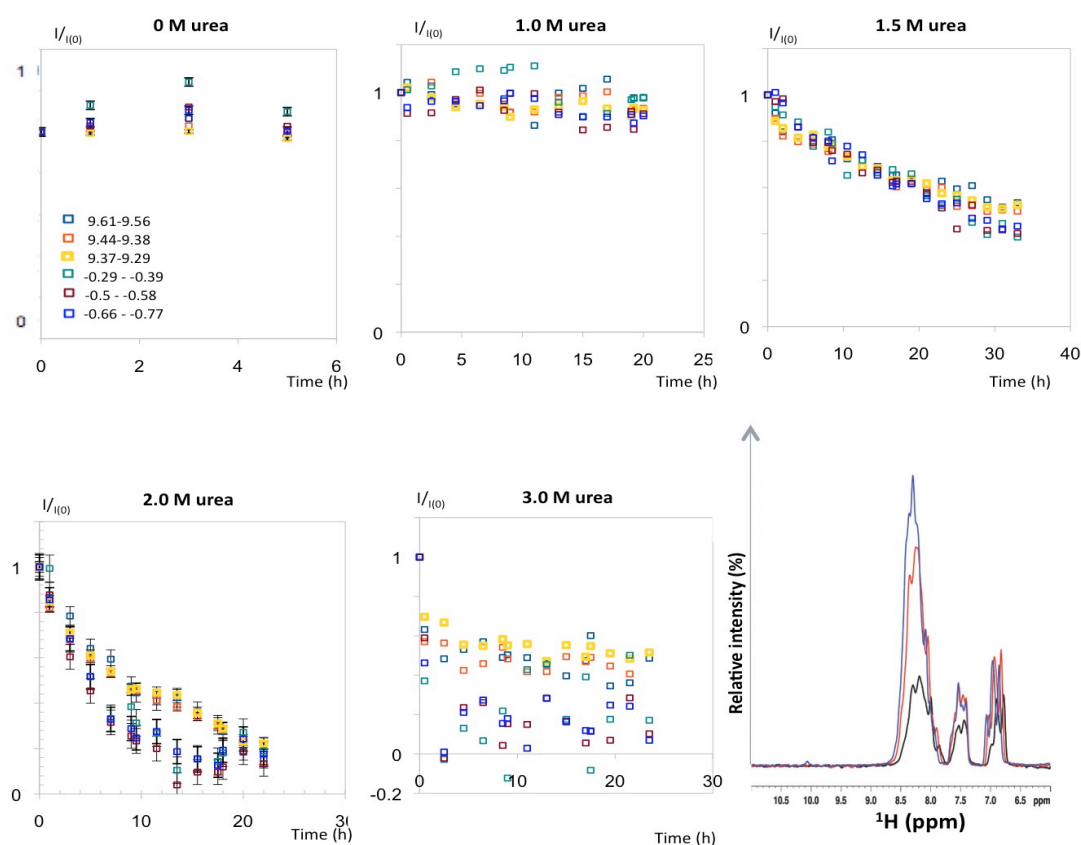


Figure 3.4. 1D ^1H -NMR spectroscopy urea titration for recombinant α_1 -antitrypsin. ^1H NMR spectroscopy spectra were recorded across the titration at 0 M, 1.0 M, 1.5 M, 2.0 M, and 2.5 M urea. In order to monitor urea induced polymerisation, intensities were measured at 3 regions upfield (9.61-9.56, 9.44-9.36, 9.37-9.29 ppm) and downfield (0.29-0.39, 0.5-0.58, 0.66-0.77 ppm). At higher urea concentrations 1D spectra supported similarly rapid equilibration to that seen at 0 M and 1.0 M urea. Lower right panel shows relative intensities for these conditions in terms of 1D projections of the 2D spectra onto the ^1H axis (black: 4.0 M; red: 6.0 M; blue: 7.6 M urea).

The upfield region is characterised by amide proton signals. While this signal would generally remain unaffected by the unfolding of α_1 -antitrypsin induced by increasing concentrations of urea, the water suppression necessary to obtain 1D NMR removes the magnetization of the water molecules of the solvent. Therefore, as the protein unfolds the NMR invisible protons of the water molecules exchange with α_1 -antitrypsin and the amide signal decays as the molecule unfolds. The downfield region is characterised by the methyl signal between 0.3 and 0.75 ppm, which is unaffected by the folding state of the protein, and generally is a great indicator of the sample concentration and integrity. The region located below zero (-0.3 - -0.8 ppm, figure 3.5), is an indicator of the folding state of the protein, where a well-folded protein lead to the appearance of small peaks in this particular region.

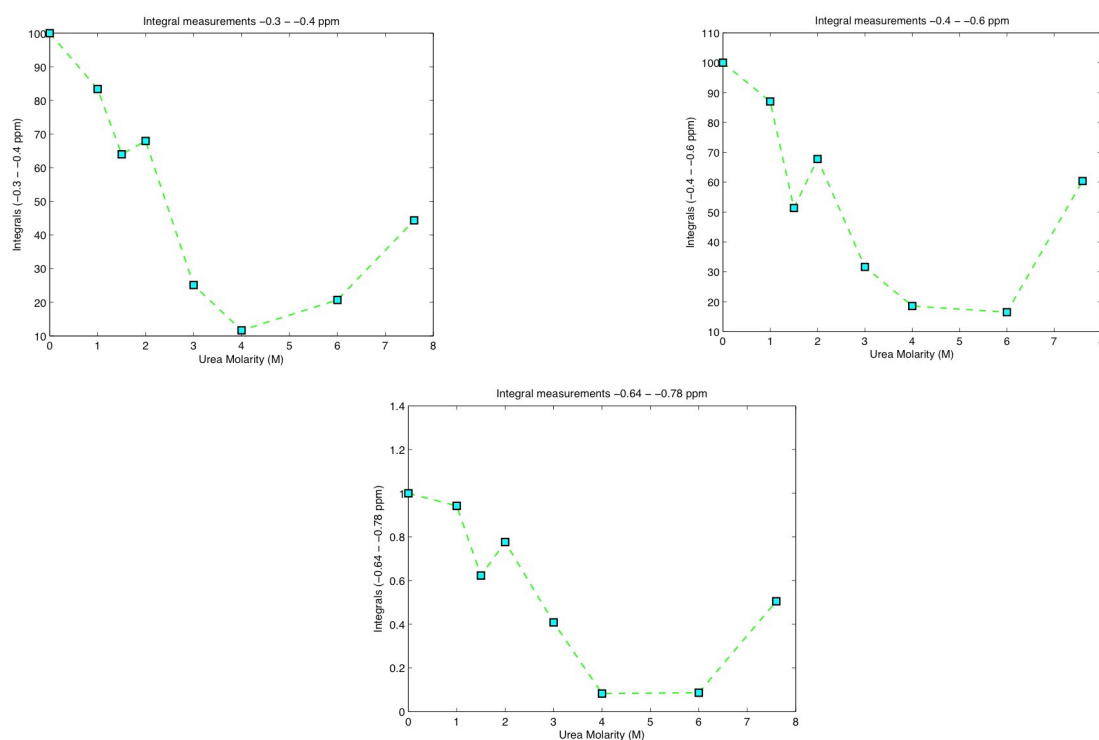


Figure 3.5. 1D ^1H -NMR spectroscopy urea titration for recombinant α_1 -antitrypsin integrals measurements. ^1H NMR spectroscopy spectra were recorded across the titration from 0M, to 8.0 M urea. In order to monitor urea induced loss of folding, integrals were measured at 3 regions below 0 ppm (-0.3 - -0.4 ppm; -0.4 - -0.6 ppm; -0.64 - -0.78 ppm).

Compared with this, the spectrum collected at 1.0 M urea reported changes in the solution behaviour of α_1 -antitrypsin (Fig.3.4.), despite the lack of unfolding in this range noted in previous studies. These therefore likely report changes in dynamic behaviour rather than a clear structural transition.

Residues in the periphery of β -sheet A reported increased conformational lability compared to their behaviour at 0 M urea. Intriguingly however, a few residues in β -sheet C reported enhanced “native-like” signal in these conditions, raising the possibility that these α_1 -antitrypsin residues are more conformationally labile at 0 M than at 1.0 M urea.

At conditions > 1.5 M urea, equilibration/pseudo-equilibration was associated with signals of polymerisation and a highly unfolded state (Fig.3.6.). Polymerisation was reported by decline in 1D NMR spectral intensity over time within a sample and/or relative to the signal intensities observed at 0 M urea, across the entire range where different polypeptide-related signals are observed. It appeared maximal at 3.0-4.0 M and minimal by 7.6 M. Unfoldedness was defined as a 2D spectral pattern of cross-peak chemical shifts closely related to those observed at 7.6 M urea.

At 2.0 M urea, assignable cross-peaks reporting on the native state were gradually lost over 7 hours and this could be detected over sequential scans. The rates of signal loss over this period varied between different residues (Fig. 3.7.). This appeared not to be simply a function of differential signal:noise ratios since cross-peaks with similar initial intensities demonstrated different decay patterns.

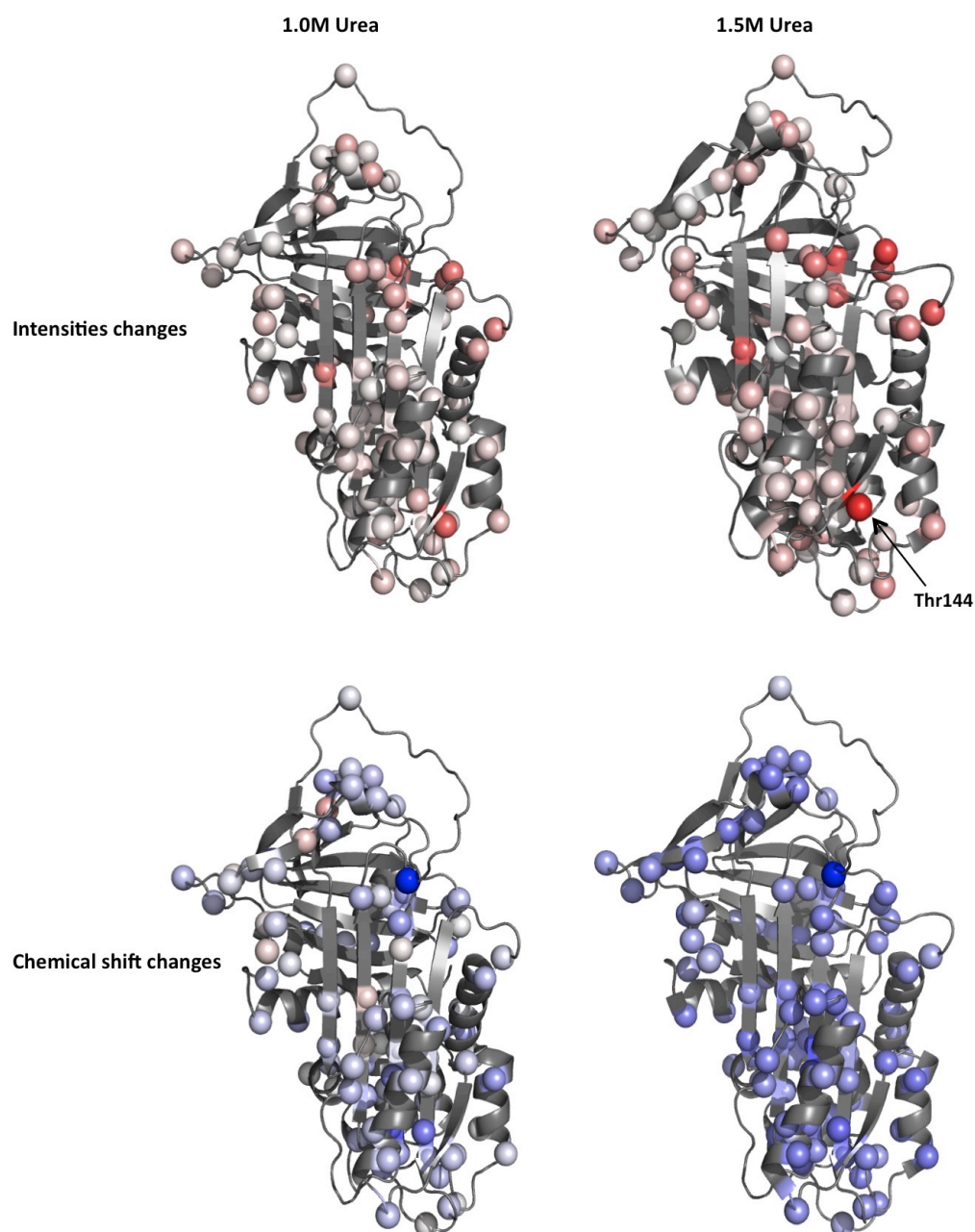


Figure 3.6. Intensity and chemical shift changes occurring at 1.0 and 1.5 M urea, residue-specific data for intensity changes. These are calculated as the ratios of intensities recorded at 1.0 M urea to those recorded at 0 M urea. Heatmap colouring here is such that increasing intensity is indicated by increasing redness and decreasing intensity by increasing blueness. The range for I_1/I_0 values is 0.00 - 1.29. No change (i.e. ratio of 1.00) is shown in white and the minimal observed value (0.00) in blue. Maximal redness therefore corresponds to a theoretical maximum ratio of 2.00. Magnitude of chemical shift changes ($\Delta\delta$): The data are represented by heatmap colouring. Increasing $\Delta\delta$ is indicated by increasing redness. This is scaled relative to all changes reported between 0 and 2.0 M urea by the subset of assigned and well-dispersed residues unambiguously observable in ^1H - ^{15}N TROSY-HSQC spectra (spheres). The colour range was therefore between no change (0.0 Hz - white) and maximal change (80.7 Hz, observed for Thr144 at 1.5 M urea - red). Residues that could not be followed in this way via ^1H - ^{15}N TROSY HSQC spectra are coloured grey.

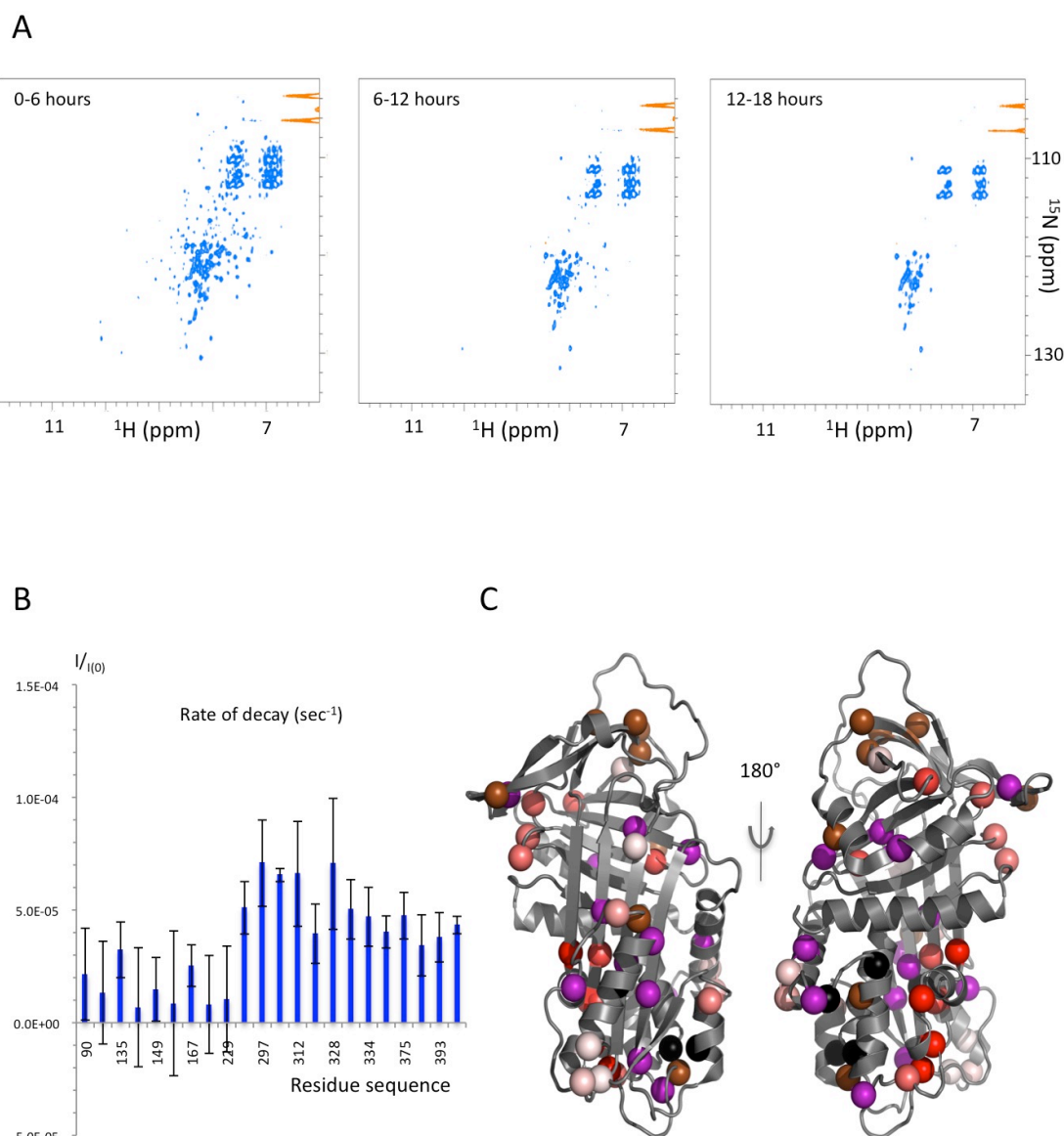


Figure 3.7. Real-time 2D NMR and intensities changes occurring at 2.0M urea.

A) ^1H - ^{15}N - TROSY HSQC NMR spectroscopy spectra were recorded across the titration at 2.0 M urea successively for 2 hours acquisition times. The spectra were grouped three at the time, 0-6 hours, 6-12 hours, 12-18 hours experiments. B) Where residues lost native-like signal over a longer period (i.e. ≥ 4 timepoints) in a progressive manner, these declines were fitted to an exponential decay function and the rate constant calculated were plotted against the amino-acids sequence; (C) These rates are indicated by colour-coded spheres representation. Residues for which signal intensity was lost at the first (0 h), second (2 h) and third (4 h) timepoints are indicated by black, brown and purple colouring respectively. The spheres with rate constants are illustrated on a red-white heatmap scale where white would represent no change over this time period from native signal intensity and maximal redness corresponds to a decay rate of 7.10^{-5} s^{-1} . Overall therefore, apparent rates of unfolding for the various residues follow the order: black>brown>purple>red>white.

The observable 2D spectra at 3.0, 4.0 and 6.0 M urea were of lower intensity corresponding to the degree of polymerisation to that seen at 7.6 M, but otherwise closely related to it. Successive urea increments from 4.0 M resulted in increased spectral intensities and gradual, small-scale migration of cross-peaks between successive conditions (Fig. 3.8).

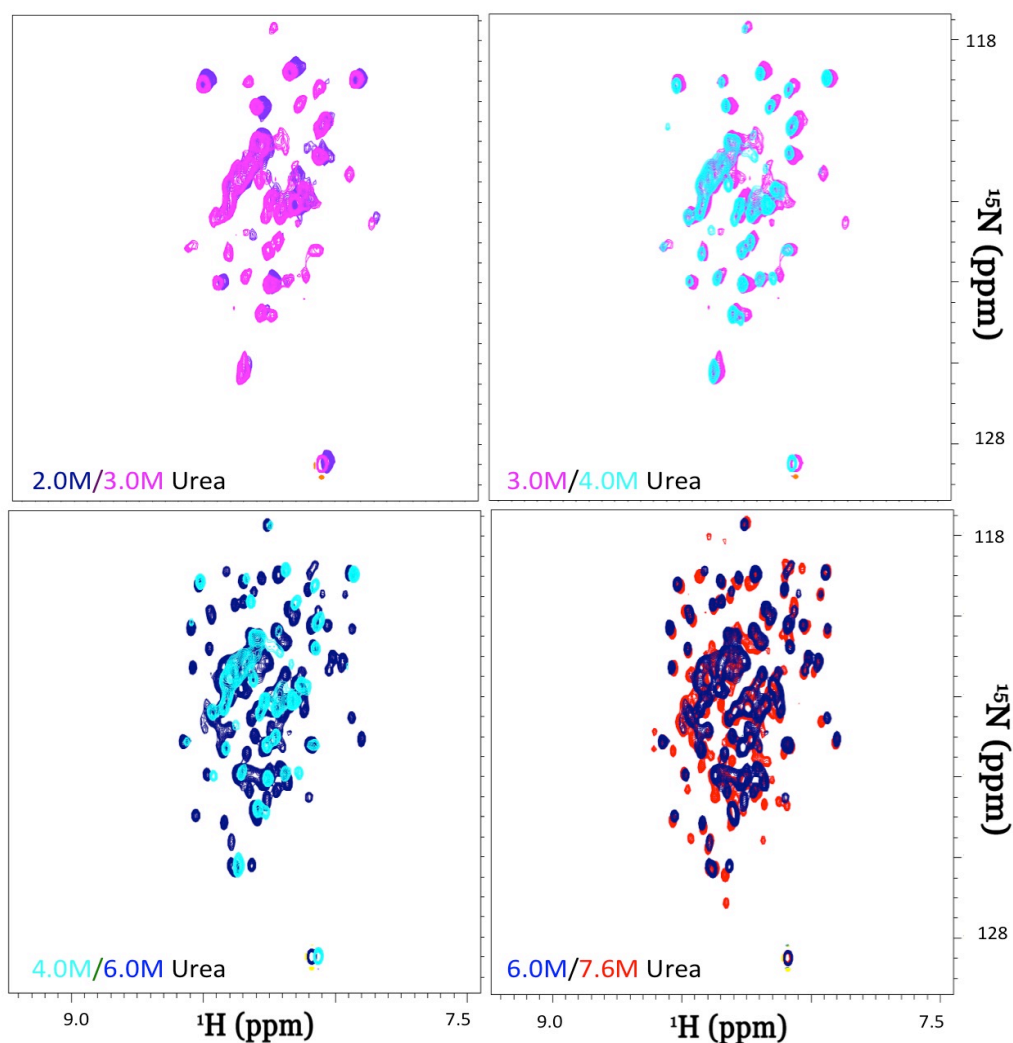


Figure 3.8. ^1H - ^{15}N -TROSY HSQC NMR spectroscopy urea titration for recombinant α_1 -antitrypsin - unfolded region. Overlays of the unfolded region of the spectra post-equilibration at 2.0 M (purple) versus 3.0 M (pink); 3.0 M (pink) versus 4.0 M (cyan) urea; 4.0 M (cyan) versus 6.0 M (blue) urea; and 6.0 M (blue) versus 7.6 M (red) urea.

The observable 2D spectra at 4.0 and 7.6 M urea were of lower intensity corresponding to the degree of polymerisation to that seen at 7.6 M, relative to the overall intensity obtained at 0.0 M urea,. This explains the expected disappearance of the well-folded peaks and the partial appearance of the unfolded signals at 4.0 M urea compared to 7.6 M urea (Fig.3.9).

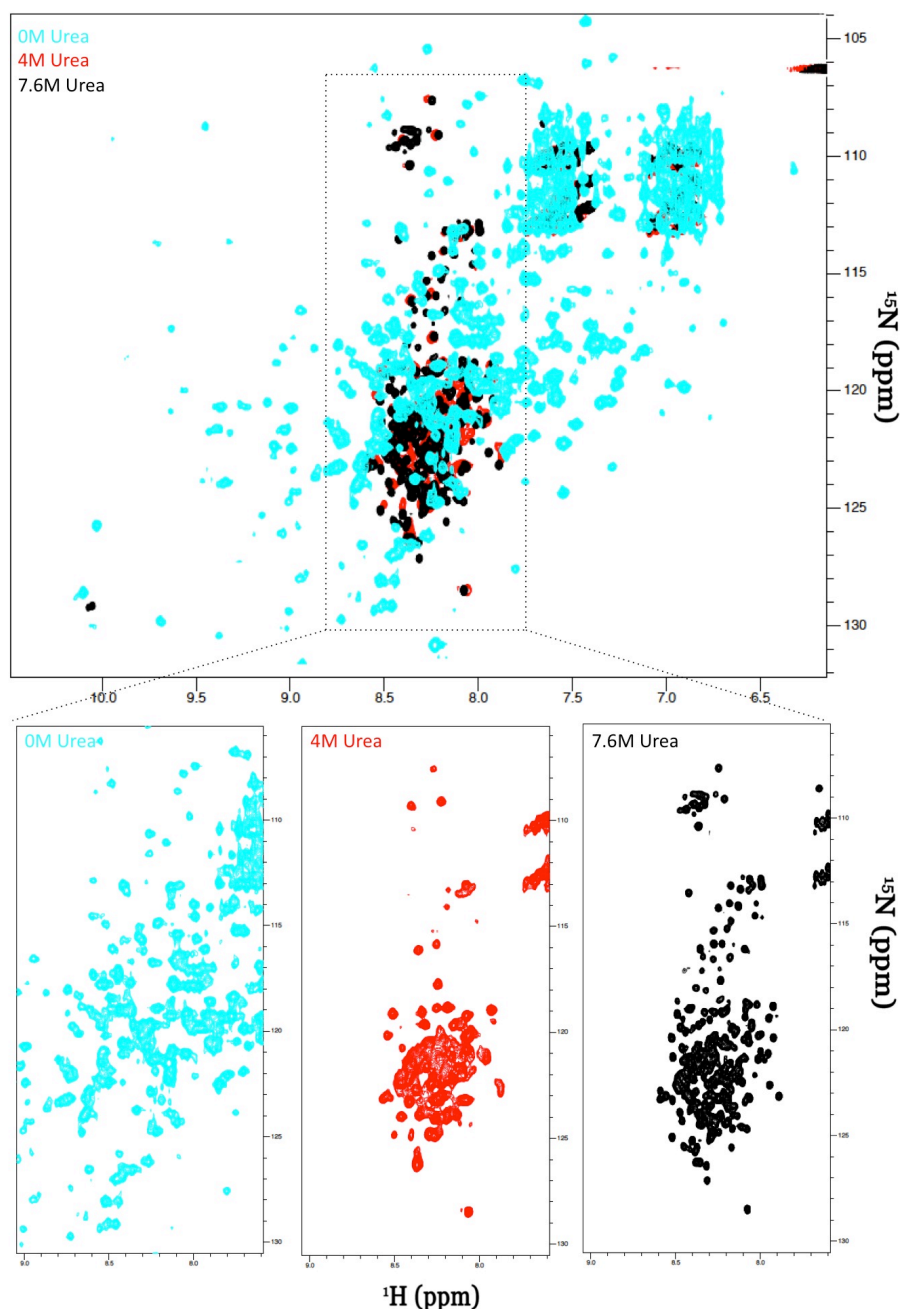


Figure 3.9. ^1H - ^{15}N -TROSY HSQC NMR spectroscopy urea titration for recombinant α_1 -antitrypsin. Overlays of 0.0 M (blue), 4.0 M (red), and 7.6 M (black) urea, and below individual spectra zoomed in between 7.5-9.0 ppm (^1H dimension) and 100-130 ppm (^{15}N dimension).

3.2.2. Circular Dichroism studies of the plasma derived and recombinant α_1 -antitrypsin

[work carried out in collaboration with Dr James Irving and Dr Imran Haq (Lomas group, CIMR, University of Cambridge)]

The absence of any NMR spectral signal clearly attributable to a thermodynamically-stabilised intermediate state in the equilibria and pseudo-equilibria studied raised the question of whether any such states were in fact present in the equilibrium state. A possible explanation for their absence was that, at the high α_1 -antitrypsin concentrations required for 2D NMR spectroscopy, intermediates were all rapidly taken up into polymers. To assess this further I undertook CD spectroscopy studies at very low α_1 -antitrypsin concentration (0.1 mg/ml, 2.2 μ M) to minimize the effects of urea-induced polymerisation while preserving changes within the fold. The low protein concentration led to relatively noisy but interpretable traces (Fig.3.10.). In order to assess in detail the secondary structure characteristics of intermediate formation I collected all interpretable far u.v. spectra for all conditions in contrast to the single (222.0 nm) wavelength studies reported previously. Surprisingly however, across urea titrations, all traces for both recombinant, non-glycosylated and plasma-derived, glycosylated (M) wild-type α_1 -antitrypsin appeared to pass through a common iso-elliptic (isosbestic) point (\sim 208 nm). Since this had to be judged for an inherently low signal:noise, the residual differences between the apparent iso-elliptic point and individual traces across the urea range were calculated. No systematic pattern of variation was discernible by this approach and so the variation was felt most likely to be a function of the noise of the system.

This type of pattern is most consistent with a two-state (native-unfolded) transition for α_1 -antitrypsin in urea. In a two-state system, in equilibrium conditions a polypeptide chain populates only native or unfolded states. The proportions of each are determined by the concentration of denaturant, but for all ratios the net CD spectral trace must pass through the iso-elliptic point that the traces associated with the folded and random coil states will share.

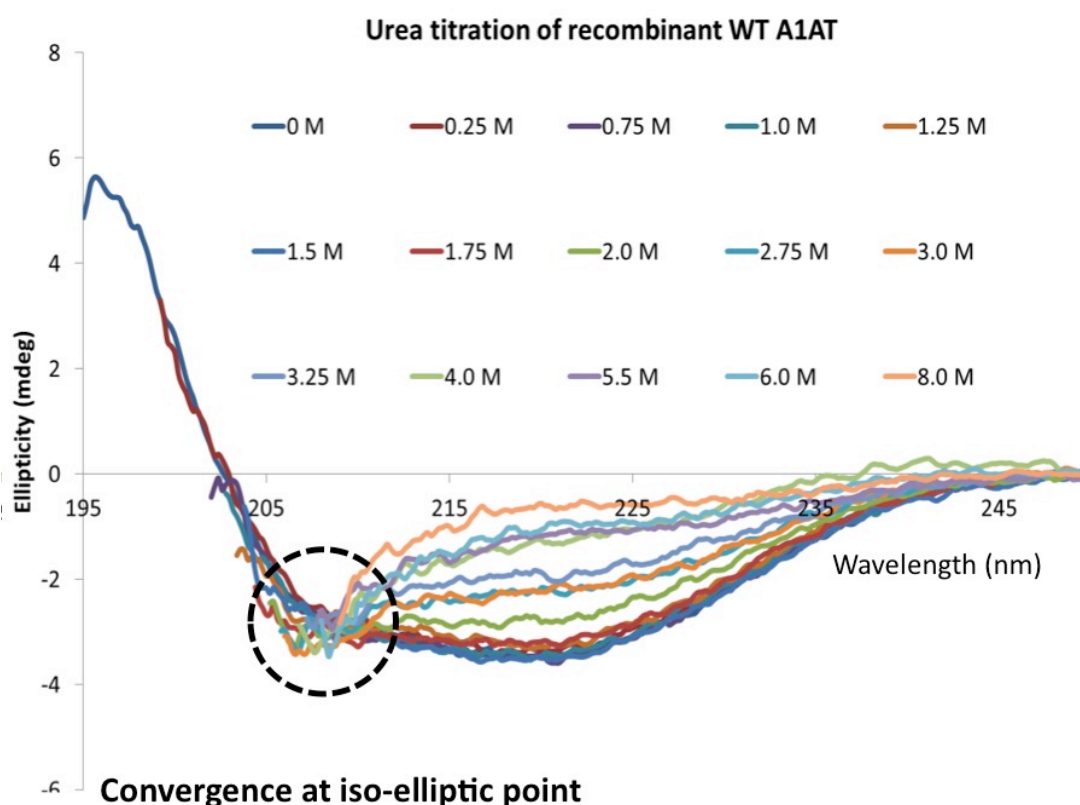


Figure 3.10. Circular Dichroism urea titration of recombinant α_1 -antitrypsin. Recombinant non-glycosylated WT α_1 -antitrypsin demonstrated unfolding behaviour converging at a single iso-elliptic point characteristic of a two-states unfolding pathway.

The apparent fraction unfolded (F_{app}) can therefore be assessed by the equation $F_{app} = (\theta - \theta_{0M}) / (\theta_{8M} - \theta_{0M})$ where θ is the ellipticity observed at a given wavelength and θ_{0M} and θ_{8M} are the ellipticities recorded at the same wavelength at 0M and 8.0M urea respectively (Fig. 3.11). All F_{app} values in the 210-230 nm wavelength ranges where folded and unfolded spectra were well diverged, were calculated. Mean (\pm /-SD) values are plotted for all variants across the urea titration.

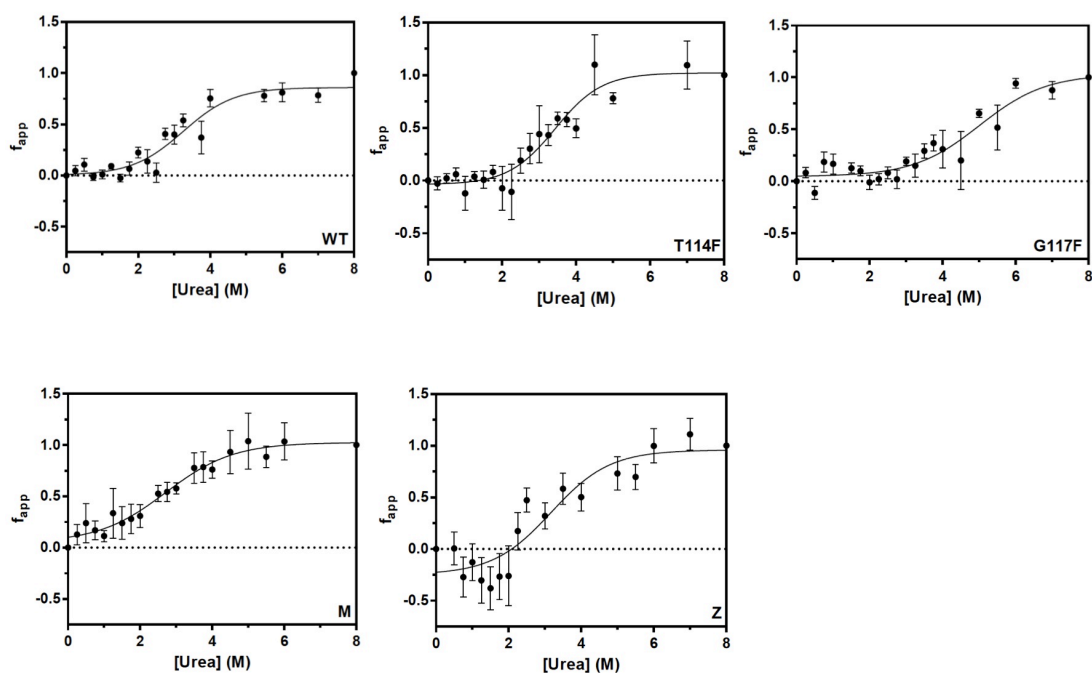


Figure 3.11. The apparent fraction folded (F_{app}) Circular Dichroism urea titration. For the plasma glycosylated M- and Z-; and recombinant WT, Thr114Phe, and Gly117Phe mutants α_1 -antitrypsin measured at 210-230 nm.

F_{app} were measured by the equation $F_{app} = (\theta - \theta_{0M}) / (\theta_{8M} - \theta_{0M})$ where θ is the ellipticity observed at a given wavelength and θ_{0M} and θ_{8M} are the ellipticities recorded at the same wavelength at 0 M and 8.0 M urea respectively.

Overall recombinant non-glycosylated wild-type and plasma derived glycosylated M α_1 -antitrypsin showed very similar secondary structure behaviour across the equilibrium state urea titration. The subtle differences that are noted may be due to glycosylation status or presence of a non-cleavable hexahistidine N-terminal tag on the recombinant constructs.

Having defined this system for wild-type α_1 -antitrypsin in the absence and presence of glycosylation I then used it to investigate the effects of polymerogenic (Z) and stabilizing (Thr114Phe, Gly117Phe) variants on the equilibrium behaviour of the protein fold in urea. Plasma-derived Z α_1 -antitrypsin was provided by kind gift of Dr Imran Haq (Lomas group, University of Cambridge) whilst the two phenylalanine variants had previously been made recombinantly and characterized within the Gooptu group (ISMB/Birkbeck,

University of London) as described previously⁸⁸. CD spectral data were again consistent with two-state behaviour for both variants and analysed in terms of the unfolded fraction (Fig.3.10.). Whilst the Thr114Phe and Gly117Phe mutation resulted in a moderate and large increases in the thermodynamic stability of α_1 -antitrypsin respectively, the Z mutation was associated with no clear reduction in thermodynamic stability as reflected by the transition midpoint ($U_{0.5}$, Table 3.1). If anything it appears to be associated with a somewhat increased transition midpoint. However the CD spectra for the Z variant (and to some degree plasma-derived M α_1 -antitrypsin that is its appropriate control) were more affected by noise than the other traces. More studies may therefore be required to confirm this surprising finding and at present the most conservative interpretation is favoured.

	WT	Thr114Phe	Gly117Phe	M	Z
$U_{0.5}$ (SD)	3.3 (0.2) M	3.4 (0.2) M	5.0 (0.4) M	2.7 (0.2) M	3.2 (0.4) M

Table 3.1. Thermodynamic stability as reflected by the transition midpoint $U_{0.5}$

3.2.3. Transiently populated α_1 -antitrypsin intermediates form off-pathway polymers with defined secondary structure in urea

The off-pathway, urea-induced polymers arose within a two-state unfolding system. Since polymers do not arise when the native state is maximally populated, this finding can only be explained in two ways. Urea-induced polymers could be formed from mechanistic but not thermodynamically stabilized intermediate species that are transiently populated in the native-unfolded transition. Alternatively, urea-induced polymers might form from unfolded α_1 -antitrypsin but only be stable at ~2.0-4.0 M urea. Indeed, unlike heat-induced or pathological polymers, urea-induced polymers do not appear to be stable at high urea concentration. However urea-induced oligomers appear sufficiently ordered to form electrophoretically distinct assemblies on TUG-PAGE (Fig.3.12). Aggregation of fully unfolded polypeptide species is best characterized in terms of the formation of disordered assemblies or amyloid species made up of many subunits. It therefore seems most likely that urea-

induced polymers form from partially folded, transiently-populated states on the urea-induced unfolding pathway.

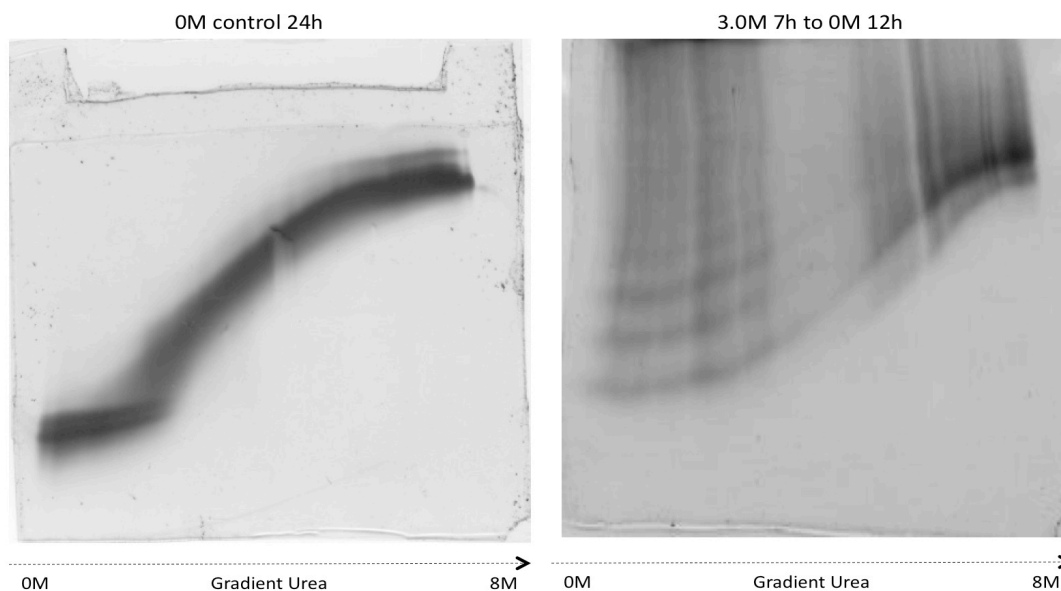


Figure 3.12. Transiently populated α_1 -antitrypsin intermediates form off-pathway polymers with defined secondary structure in urea. Transverse urea gel PAGE (TUG-PAGE) Samples of α_1 -antitrypsin at $\sim 1\text{mg/ml}$ were subjected to non-denaturing conditions in 0-8M Urea gradient formed within a 7.5 % (v/v) polyacrylamide gel at A- 0 M, and B- 3.0M urea. Urea-induced α_1 -antitrypsin polymers ($100\text{ }\mu\text{M}$ α_1 -antitrypsin, incubated for 7 h at 3.0 M urea and then incubated at 0M for 12h to recover any potential non-irreversible conformer formation) do not show the stability at high urea concentrations on TUG-PAGE that characterises heat-induced or ex vivo polymers. However they do show discrete oligomeric behaviour at low urea concentrations.

3.3. Discussion

It is important to characterize the folding/unfolding behaviour of α_1 -antitrypsin since various lines of study have indicated that intermediates along these pathways are responsible for its polymerisation in disease. The biggest burden of α_1 -antitrypsin polymer load appears to be within its folding compartment: the hepatocyte ER¹⁰². The Z mutation has been proposed to perturb normal folding of α_1 -antitrypsin¹⁷¹ and thermodynamically destabilize its native conformer¹⁰³. Various biophysical data have supported the presence of thermodynamically stabilized intermediates across denaturant titrations^{81; 83} and upon heating^{101; 113}. It is also a goal of the host groups for this project to use NMR spectroscopy and complementary methods to understand folding of the polypeptide nascent chain on the ribosome and within the ER, and also how pathogenic mutations affect this. Urea equilibrium titration studies are the classical way to explore the folding behaviour of a full-length polypeptide chain by NMR spectroscopy¹⁷². Such studies were therefore undertaken for α_1 -antitrypsin.

The NMR spectroscopic data in this chapter identify residue-specific changes in native wild-type α_1 -antitrypsin at low urea concentrations (1.0 M urea). These likely reflect alterations in conformational dynamics (kinetic instability) rather than a transition to a different state. A few residues, e.g. in β -sheet C show this as a reduction rather than an increase in native-like signal, and may indicate features that become more stable and so prevent global unfolding in low denaturant conditions. At 1.5 M urea the changes become more generalized and are compounded by a low constant level of sample loss due to polymerisation.

In the urea range 2.0-6.0 M, with an α_1 -antitrypsin concentration of 100 μ M, urea-induced polymerisation becomes the dominant phenomenon. Residual monomeric polypeptide signal at equilibrium is highly similar to that of the fully unfolded state maximally populated at 7.6 M urea. Despite the occurrence of polymerisation, no signal attributable to a distinct thermodynamically-stabilised

intermediate state is observed. It seemed likely that, at relatively high α_1 -antitrypsin concentrations, any such monomeric intermediate could be rapidly taken up within polymers. To explore this possibility, CD spectroscopy was used for complementary urea equilibrium studies at 50-fold lower protein concentration where polymerisation effects would be much less evident. This system also allowed relatively straightforward assessment of, and comparison between: wild-type, stabilized and disease-variant recombinant (non-glycosylated) and plasma-derived (glycosylated) α_1 -antitrypsin.

Remarkably these data also strongly supported the 2 state solution behaviour of α_1 -antitrypsin in urea. These data are in apparent contradiction to earlier urea equilibrium studies using more limited CD spectral findings that defined an intermediate state based upon a plateau phase around 4.0 M urea within the loss of signal associated with unfolding^{83;166;81; 100; 167}. The discrepancy may best be explained by confounding effects of urea-induced polymerisation at the higher α_1 -antitrypsin used in the earlier work. The absence of an intermediate species characterised by urea equilibrium studies does not necessarily suggest that intermediate states cannot be accessed in other unfolding (e.g. guanidine) systems. Moreover the absence of a thermodynamically stabilised intermediate in equilibrium conditions does not preclude the existence of a kinetic intermediate. Similarly it does not preclude the potential for intermediates along other polymerogenic pathways within the energy landscape (as illustrated in fig.1.16.). Such a pathway will be explored in Chapter 5 in the context of a deficiency mutation.

In addition the relatively mild thermodynamic destabilization of the native fold by the Z mutation in plasma-derived α_1 -antitrypsin contrasts with its highly polymerogenic effects in disease and when native protein is incubated at physiological temperatures¹⁴⁸. This is demonstrated by the opposite effects of the 'stabilizing' Thr114Phe and Gly117Phe mutations. They increase the thermodynamic stability of α_1 -antitrypsin more markedly than the Z mutation reduces it. However they have no rescue and minor rescue effects, respectively, on ER polymer load and native α_1 -antitrypsin secretion when introduced on a Z

background in eukaryotic (including mammalian) cell models¹⁷⁰. Assuming that the thermodynamic effects of distant point mutations should be additive, the data in this chapter suggest that the major disease-relevant effects of the Z mutation arise through kinetic rather than thermodynamic destabilization. Conceptually this may equate to an increase in native state dynamics.

This chapter's findings agree well with the characterization of the effects of the Z mutation on recombinant α_1 -antitrypsin produced in a yeast system¹⁷¹ and therefore resolves the apparent contradiction between these data and previous thermodynamic studies of plasma-derived protein. In the latter¹⁰³, the presence of the Z mutation was apparently associated with major thermodynamic destabilization that correlated closely with polymerogenicity. However, I and my colleagues have recently shown that the assay used for these studies (thermal ramp assessed by CD spectroscopic ellipticity at 222 nm) reports a composite of unfolding and polymerisation events¹⁷³. The 0.5 mg/ml α_1 -antitrypsin concentration used by Dafforn and colleagues to characterize the plasma-derived protein is much higher than that used to characterize the recombinant, non-glycosylated form. It is therefore most likely that the apparently close inverse correlation between thermodynamic stability and polymerogenicity reported was in fact also confounded by polymerisation effects in the stability assay.

**Chapter 4 – STRUCTURAL DYNAMICS OF THE NATIVE α_1 -
ANTITRYPSIN STUDIED BY HYDROGEN DEUTERIUM EXCHANGE
NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY**

4.1. Introduction

4.1.1. General principle of HDXNMR

Members of the serpin superfamily of proteins regulate key physiological processes through their ability to undergo major conformational transitions¹⁷⁴. As was described in Chapter I, function is intimately related to the unusual metastability of the native state to which serpins fold. Pathogenic mutations trigger abnormal conformational transitions to populate intermediate state(s) that result in serpin polymerisation through a mechanism that remains controversial^{96; 97; 99}. Using NMR spectroscopy, we are investigating the interrelated processes of serpin folding, misfolding and polymerisation in solution using the 45 kDa prototypic serpin α_1 -antitrypsin. The recent assignment of the backbone resonances of α_1 -antitrypsin by our group has been a major breakthrough in this regard¹⁷³. It allows us to ask more sophisticated questions by a range of NMR techniques, for instance to understand how the dynamic properties of individual residues within native α_1 -antitrypsin may facilitate conformational change.

To probe the conformational dynamics of the native state in solution over slow (>ms) timescales, backbone amide hydrogen/deuterium (HD) using NMR (HDXNMR) exchange was used. Figure 4.1 shows the HDXNMR method based upon a chemical reaction in which a covalently bonded hydrogen amide atom is exchanged by a deuterium atom. The rate of exchange is related to the degree of solvent exposure and protection due to secondary structure interactions within a molecule. HD exchange can be monitored by Mass Spectrometry (HDXMS)^{90; 175; 176; 177; 178} at the level of peptide fragments, or at the residue-specific level using NMR spectroscopy¹²⁹. The rate of exchange for protein backbone amide hydrogen atoms varies across timescales from seconds to years. The exchange rate is slow when the amide is buried in the core of the protein, or when it is involved in a hydrogen bond (closed)¹⁷⁹. Conversely exchange will be faster when it is exposed to the solvent (open) (Fig.4.1.A). HDXNMR can be monitored by recording the stability or disappearance of signal over time after rapid solvent

exchange into D₂O using ¹⁵N-¹H-correlation NMR experiments, typically ‘rapid acquisition’ SOFAST-HMQC experiments¹³⁸ are used.

Solvent exchange is undertaken using lyophilization, a desalting column, or dilution into D₂O. The opposite process (i.e. gain of signal) occurs if a deuterated sample is solvent exchanged into H₂O. Here, the former approach has been used. As for the concepts underpinning the experiment, the Linderstrøm-Lang model of H-D exchange proposes that the rate of H-D exchange: k_{ex} , can be related to the open/closing of structural segments of the protein¹⁸⁰. In the ‘closed’ state, the amides are protected and less likely to exchange with the deuterium of the solvent. In the ‘open’ state, the proton will exchange with an intrinsic rate, k_{int} , that depends upon pH, temperature and the protection provided by the peptide bond and the side-chains of the two adjacent amino-acids. The relationships of k_{int} with these variables are known.

Therefore, by comparing the observed exchange rate k_{ex} of a particular backbone amide proton within the sequence at a specific temperature and pH with k_{int} , the degree of protection afforded by local structure and dynamics can be quantified. k_{ex} can be derived by fitting intensity decay data of the amide backbone resonance to the exponential function: $I(t) = I_0 \cdot e^{-k_{ex} \cdot t} + c$ (Fig.4.1.B), where $I(t)$ is the intensity at time t following solvent exchange and I_0 is the initial intensity before solvent exchange. Once the constant of exchange is characterised at a residue-specific level, two regimes can be analyzed depending on the relative values of k_{int} and k_{cl} .

Since proton exchange will be rapid at high pH and slow at low pH, k_{ex} values relate to pH (Fig.4.1.C) with the relationship determined by the k_{int} and the protection factor, P for the individual residue. In all cases the curve is characterised by complete pH independence at high pH (where the rate-limiting step is achieving open state, $k_{op} = k_{ex}$, $k_{int} \gg k_{cl}$: EX1 exchange regime) and complete pH dependence at low pH values (limiting step is HD exchange, EX2 regime). In the EX2 regime, k_{ex} is directly related to the protection factor, P (Fig.4.1.C). The determination of the protection factor for each amino acid of a protein can be used to probe its stability, folding or analogous conformational

change. Therefore, under appropriate conditions, the labile hydrogen atoms in a region of a folded protein that unfold cooperatively will be exchanged simultaneously and irreversibly at all solvent exposed sites. P relates to the difference in free energy of the observed residue, $\Delta G_{\text{op-cl}} = RT \ln(P)$ between the “closed” state and the “open”, exchange-competent state, which also corresponds to $\Delta G_{\text{app}} = -RT \ln(k_{\text{op}})$ where $k_{\text{op}} = 1/P$. By measuring the protection factors of a different folding equilibrium than the native equilibrium (using chemical denaturation for example), the folding free energy landscape can be mapped at a residue specific level^{181; 182; 183} and the protection factors of the native state can be used to map the stability of the structure^{179; 184; 185; 186}. Furthermore, the use of HDX-NMR also provides insights into intermediates populated in solution^{120; 186; 187} at equilibrium.

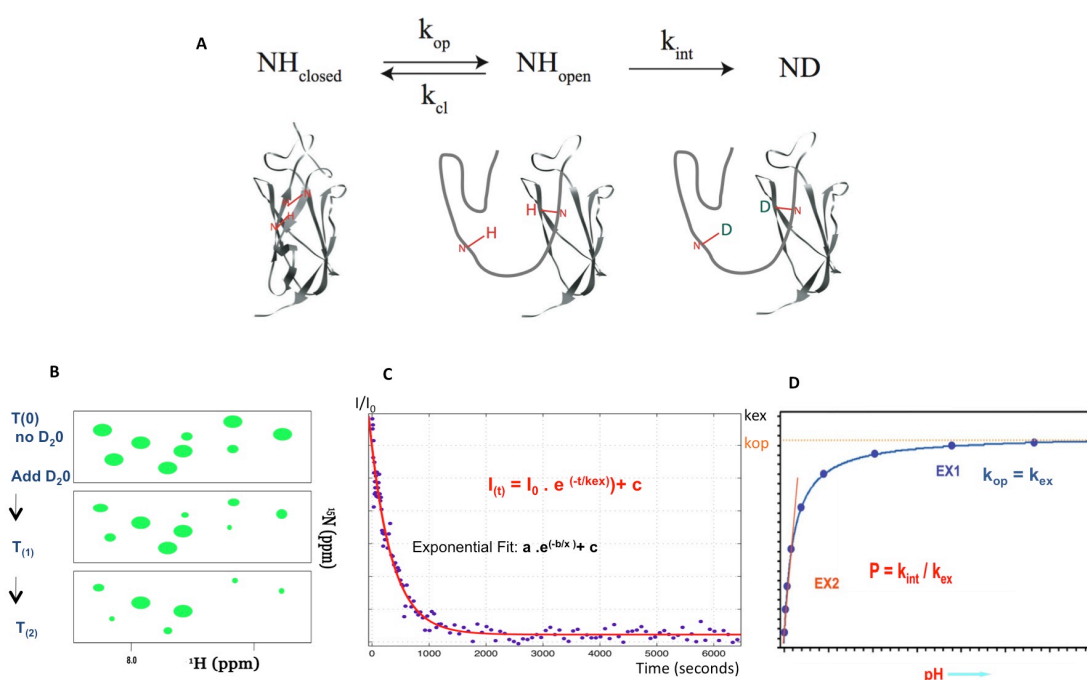


Figure 4.1. Hydrogen Deuterium exchange. A- Linderstrøm-Lang model of hydrogen exchange reaction. Refer to main text for definitions of rates as well as for details of the analysis of the different HD exchange regimes¹⁸⁸. B- Schematic illustration of HDX-NMR to allow the determination of k_{ex} by fitting (Panel C) the decay in intensity at a residue-specific level to an exponential decay function using the formula indicated in red; the variation of k_{ex} across the pH range (Panel D), allows the determination of regime (EX1 or EX2), and the calculation of the protection factor, P , for residues in the EX2 phase, using the ratio $P = k_{\text{int}} / k_{\text{ex}}$ and the free energies as described in the main text.

4.1.2. HDXMS studies of α_1 -antitrypsin

HD of α_1 -antitrypsin in the native and polymeric states and during complexation with β -trypsin has previously been assessed using mass spectrometry (HDXMS)^{90; 91; 92}. When HDXMS data are obtained, resolution is limited by the lengths of the peptide products of enzymatic treatment. Nevertheless, these studies have provided a number of useful insights. Firstly, HDXMS of native α_1 -antitrypsin showed a highly dynamic F-helix, particularly at its C-terminal end⁹². In contrast, strands 3 and 5 of the β -sheet A, which must separate to allow insertion of the RCL, during both polymerisation and enzyme inhibition, are highly stable. Strand 1C (s1C), which anchors the distal end of the RCL is also stable, while strand 6 of the β -sheet A was found to be surprisingly flexible. A subsequent study investigated the conformational changes in α_1 -antitrypsin during complex formation with β -trypsin¹⁷⁶. The data indicated that a substantial proportion of α_1 -antitrypsin was unfolding transiently before refolding, not only in the expected regions, such as RCL, helix F, and following loops; but also helix A, strand 6 of the β -sheet B (s6B) and the N-terminus.

More recently, the conformational changes during heat-induced polymerisation were characterised using HDXMS by Wintrode *et al*⁹⁰. Heat-induced polymerisation was found to be preceded by a significant destabilization of β -sheet C. As expected the flexible RCL became far more strongly protected (5-10 % and above 20 %) once polymerisation had occurred, and the centre of β -sheet A and helix F became substantially more protected (10-20 % and above 20 %). These data are consistent with a single-strand model of polymerisation³⁷.

A second model of polymerisation has been derived from the crystal structure of a closed dimer of antithrombin⁹⁷ that involved a β -hairpin domain swap of the RCL and strand 5 of β -sheet A (s5A) comprising 50 residues, between molecules. The dimer and intermediate states studied biochemically in the same work were induced by low pH and/or mild denaturant (GuHCl) conditions. It was proposed that, if considerably more structure were unfolded relative to the native fold (helix I and its linker to s5A), such a β -hairpin

intermolecular linkage could also underlie the formation of extended polymers. This would also lead to the observed increase in protection of the RCL and s5A regions data. However upon polymerisation, helix I, and adjacent residues demonstrate a small increase in protection from exchange by HDXMS (5-10 %) rather than the reduction that the β -hairpin model would predict. A subsequent study showed that a monoclonal antibody (2C1) specific for α_1 -antitrypsin polymers found in hepato-cellular inclusions¹⁸⁹, reacted strongly with heat-induced polymers but not with those induced by guanidine⁹³, urea or low pH. The findings indicated that polymers produced *in vivo* are structurally distinct to those induced by partial denaturation.

The most recently proposed model arises from a low resolution structure of a self-terminating trimer of α_1 -antitrypsin (PDB structure: 3NDD)⁸⁸. This was formed by prolonged heating of an engineered mutant of α_1 -antitrypsin in which conformational lability was constrained by an artificial disulfide bond between strands 5 and 6 of β -sheet A. This species does contain the 2C1 monoclonal antibody recognition epitope that differentiates polymers found in liver disease or formed by heating from those formed in urea or guanidine⁸⁸. The linkage demonstrates intermolecular insertion of the RCL into β -sheet A. The intermolecular linkage is a domain swap of the three β -strands between the RCL and the C-terminus involving the s1C and the core strands 4 and 5 of β -sheet B. None of those motifs were observed to demonstrate a significant increase in protection upon polymerisation in the HDXMS study (0-10 %). However, the likely slight increase in protection of the three β -strands between the RCL and the C-terminus involving the s1C and the core strands 4 and 5 of β -sheet B upon polymerisation might be difficult to differentiate from that of the native state from changes in peptide deuteration rates. Therefore validation of the third model of polymerisation over another using HDXMS data is difficult.

Together, these data from HDXMS studies surveyed the lability of α_1 -antitrypsin in the native state and evaluated of changes in protection upon polymerisation and during enzyme inhibition, typically with ~70 % sequence coverage. Nevertheless, the resolution of this technique is limited by the lengths

of the tryptic-digested peptides, assessed by MS. These may report the combined effects from contiguous residues that behave differently e.g. contributing to different structural motifs. HDXNMR studies were therefore undertaken in order to characterize backbone amide protection and dynamics at the level of individual residues.

4.2. Results

4.2.1. Sample integrity

Using biochemical techniques aiming to measure the kinetics of polymerisation of α_1 -antitrypsin¹⁹⁰, Mahadeva et al., measured the rate of polymer formation over a range of pH by measuring the intrinsic tryptophan fluorescence of plasma α_1 -antitrypsin (excited at 295 nm and measured at 340 nm). Their results showed that α_1 -antitrypsin begins to undergo acid or base-induced polymerisation between pH 6.0 and 5.0, and between pH 8.0 and 9.0, respectively. The titration used in this analysis was therefore performed across the range pH 6.0 - 8.0 (equivalent to pD* 6.4 - pD* 8.4). However, in order to increase the number of residues reaching the EX2 phase, the pH 5.75 (pD* 6.25) was also monitored, and together, these experiments were coupled with a complete set of biophysical and biochemical characterisation experiments to assess this into detail.

Since pH conditions can induce polymerisation, the integrity of the protein samples used for HDXNMR analysis under different pH conditions was tested during each HDXNMR experiment using ¹H NMR. In particular cases, MS experiments and functional activity assays were also used. Figure 4.2 illustrates the characterisation of the HDXNMR samples using ¹H NMR and MS, which was undertaken for every pH condition. Since α_1 -antitrypsin begins to undergo acid-induced polymerisation between pH 6.0 and 5.0, the characterisation of the sample integrity at pH 6.0 is shown, as a typical example of the type of characterisation undertaken in these analyses.

Consideration of the ¹H 1D NMR spectrum during the HDXNMR acquisition period from t= 0 to t= 4 days (Fig.4.2.A) and in particular focusing on

the methyl region (Fig.4.2.B) can inform on sample integrity as this region is unaffected by exchange with deuterons. The excellent overlay observed indeed, shows that the entire sample has retained its native structure. In addition, post-HDXNMR all samples were characterised using MS (Fig.4.2.C.). The spectra obtained show that the proportion of monomeric species present in the sample post-NMR at pH 6.0, was identical to that observed for the control (pH 8.0, identical concentration: 250 μ M, and 100% active).

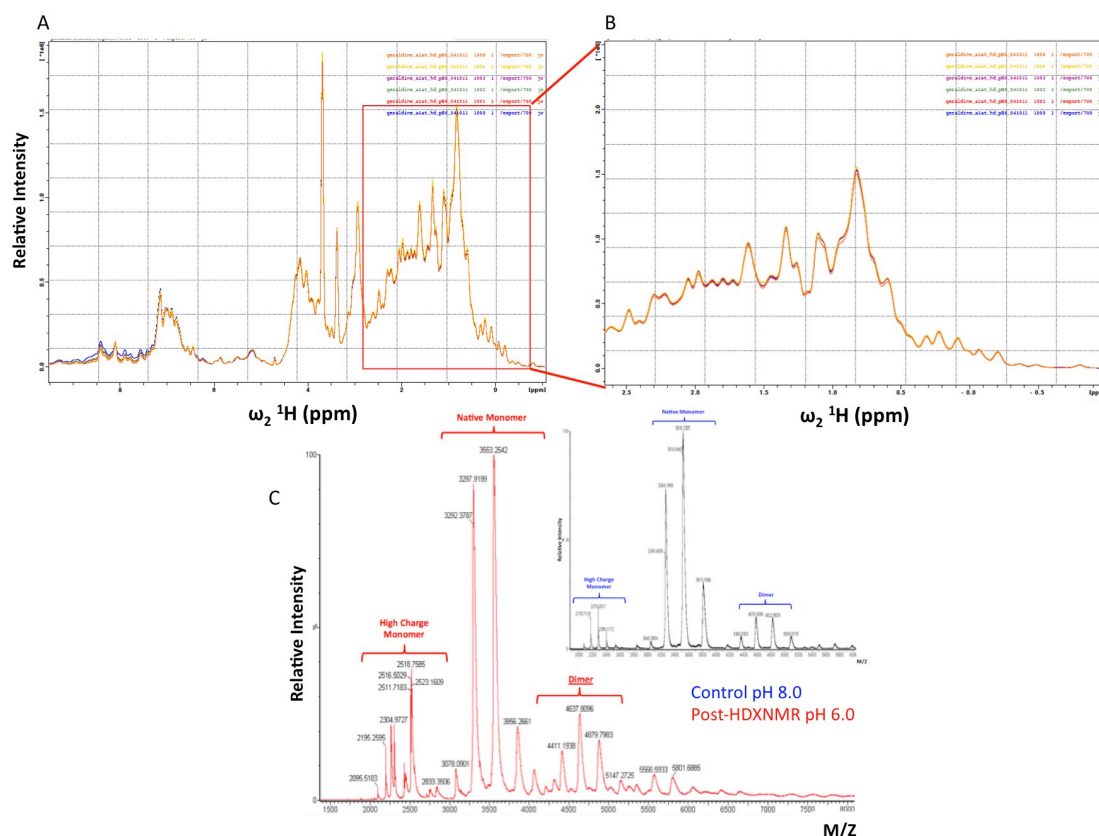


Figure 4.2. HDXNMR protein integrity characterisation using ^1H NMR and MS: example of pH 6.0 set of data from t=0 to t=4 days. A- Represents an overlay of HDXNMR ^1H NMR experiments monitored from t=0 to t=4 days at pH 6.0 across the ppm range. B- Represents a zoom-in of the same overlay of the methyl region unaffected by the exchange. The former illustrates perfect overlay between t=0 and t=4 days, demonstrating that the structure of α_1 -antitrypsin at pH 6.0 during the HDXNMR experiment remained unchanged. C- The characterisation of the sample integrity was completed with MS experiment pre- (blue) and post - (red) HDXNMR experiments. This data shows that the quantity of monomeric species present in the sample post-HDXNMR at pH 6.0 remained unchanged compared to the control (pH 8.0).

In order to increase the number of residues reaching the EX2 phase in this analysis, HDXNMR was undertaken at pH 5.75 where the exchange is expected to occur even slower than at pH 6.0. However, as described before, this pH condition corresponds to the range where α_1 -antitrypsin begins to undergo acid-induced polymerisation between pH 6.0 and 5.0. Therefore in order to monitor the integrity of the sample present in solution, ^1H 1D NMR was used (Fig.4.3.A.).

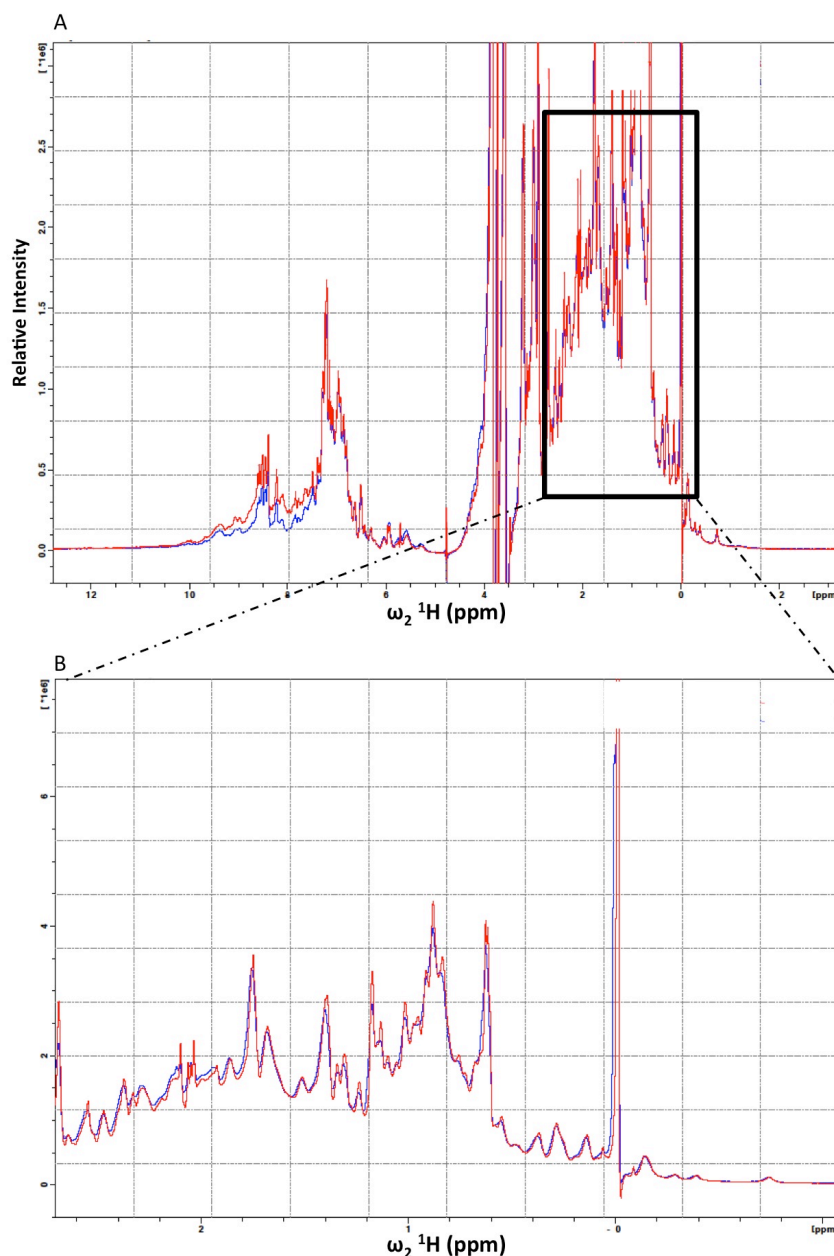


Figure 4.3. HDXNMR protein integrity characterisation using ^1H NMR for pH 5.75 from $t=0$ (red) to $t=4$ days (blue). A- Represents an overlay of HDXNMR ^1H NMR experiments monitored from $t=0$ to $t=4$ days at pH 5.75 B- Represents the zoom-in the methyl region 0.0 – 2.0 ppm.

The methyl region (0-2 ppm, Fig.4.3.B) which is characteristic of the sample integrity, and homogeneity, unaffected by the exchange occurring in the sample, was overlaid at $t = 0$ (red) and $t = 4$ days (blue). A 15 % loss in intensity in the methyl region, occurring steadily over the 4 days of HDXNMR, reported loss of detectable monomeric sample due to polymerisation occurring at a constant rate. Before analysis of HD exchange rates, the intensity data recorded at pH 5.75 were therefore scaled in a time-dependent manner to correct for sample loss using the following formula:

$$Corr (t=T_{tot}) = 1 - [\Delta_{corr} \times t/T_{tot}] \quad \text{Equation 3.5}$$

(where $\Delta_{corr} = 0.15$)

4.2.2. Reproducibility of HD exchange at pH 7.0.

Typically HDXNMR experiments record single datasets for each pH condition, and internal consistency between data at different pH-points is assumed to reflect accuracy. However in order to assess the reproducibility of the derived k_{ex} values, I undertook a repeat of HD exchange experiments at pH 7.0 and compared the observed k_{ex} values (Fig.4.4).

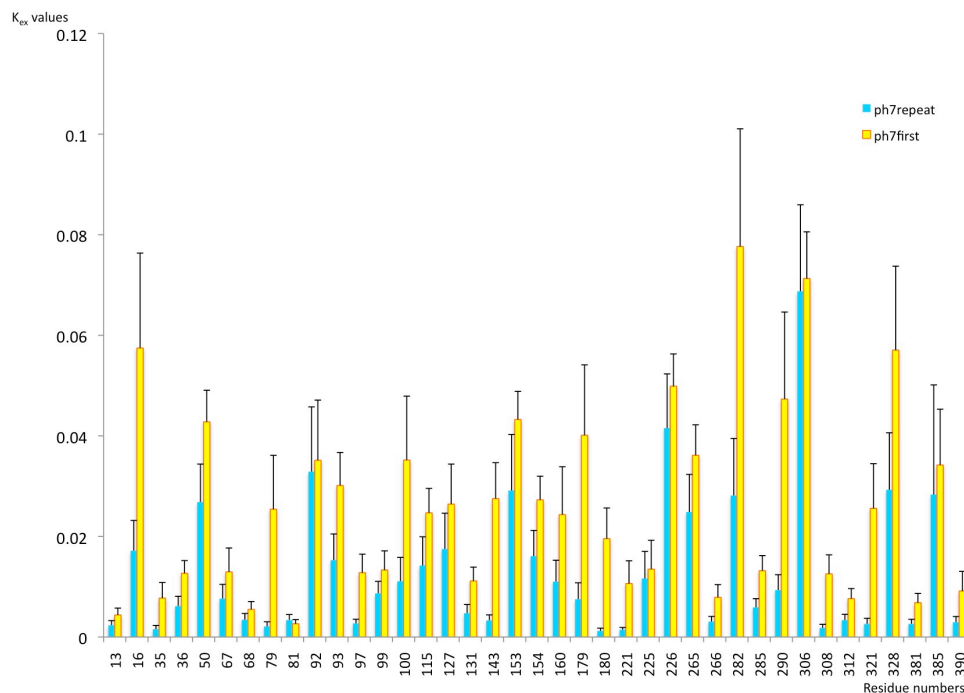


Figure 4.4. HDXNMR of α_1 -antitrypsin at pH 7.0. Comparison of the rates of backbone amide HD exchange (k_{ex}) assessed via independent experiments at pH 7.0 for residues where the curve fits allowed R^2 values > 0.8 . The green represents the first set of data monitored at pH 7.0, and the yellow corresponds to the repeat of this experiment.

The details of the HDXNMR results at pH 7.0 including the spectra and the fits are described in Section 4.2.2. (*vide infra*). I focus here initially however on the reproducibility of these fits in order to show that this method is viable. This approach showed that $\sim 2/3$ of k_{ex} values from the 2 experiments were broadly consistent (defined as overlap of error bars in Fig.4.4). Since data for a substantial minority of residues gave apparently poor agreement between experiments, I interrogated the residue-specific data further by visual analysis of data from residues that showed poor correlation between k_{ex} values. This demonstrated two reasons for such discrepancies. Data from some residues showed curve-fits that were visually poor despite high R^2 values in one of the two datasets. These were removed from the analysis. In other cases the data appeared broadly consistent but the curve-fits were best optimized for different elements of the curve. Typically these fitted the transition between initial rapid decline and curve plateau in different ways, since this was the least linear section. In these cases the situation did not resolve when data in the plateau

phase were truncated, or when data from sequential time points were averaged together in sets of four to smooth the curve. Faced with a choice between discarding data from one of a pair of similarly valid fits, or improve accuracy at the expense of consistent handling of different datasets, I opted in such cases to retain the mean of the alternative k_{ex} values.

The grey dots (Fig.4.5.) show the paired k_{ex} values retained following this analysis. The trend line obtained for the correlation was shifted towards a better agreement by $\sim 5\%$ (from $y = 0.5333x - 0.0016$, to $y = 0.5861x - 0.0023$) following this quality control step. Furthermore, the R^2 value for the correlation was also improved ($\sim 12\%$) using this strategy.

Overall the findings were reassuring that, when curve-fits were assessed by both R^2 values (≥ 0.8) and by visual inspection, the data selected were sufficiently robust to rely on single datasets across the rest of the pH titration. The remaining errors are likely inherent due to the large size of the system being studied by NMR spectroscopy. In contrast, visual evaluation of curves for which the R^2 values were 0.6 – 0.8 confirmed that the fits were insufficiently successful to have confidence in the derived k_{ex} values compared with those retained with $R^2 > 0.8$.

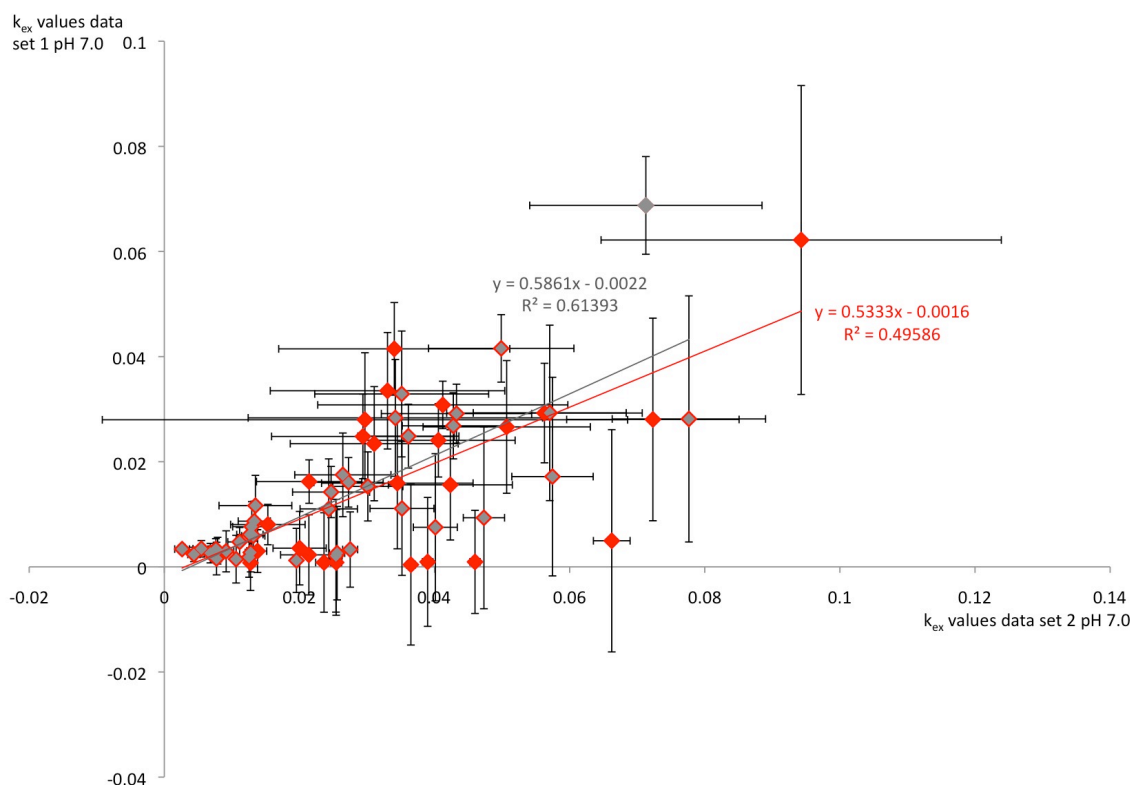


Figure 4.5. Correlation plot comparing the rates of exchange comparison between HDX-NMR sets of data monitored pH 7.0. Red dots show the relationship between the rates of exchange if all curve fits with R^2 values of ≥ 0.8 were accepted. Grey dots indicate those k_{ex} values retained from this set following visual evaluation of fit quality. For residues in this selected subset mean k_{ex} values were carried forward for comparison with those obtained at other pH points.

4.2.3. Patterns of HDX-NMR observed for α_1 -antitrypsin

HDX-NMR was undertaken at six pH points between pH 5.75 and pH 8.0, as described in the Chapter II. Typically over 150 spectra were recorded at each pH value. Spectra were typically acquired for 4 min each ($ns=8$), with experiments with increased numbers of transients ($NS = 16, 32, 64, 128, 256, 512$ for increased signal). The early, HDX-NMR experimental timepoints collected over short (e.g. 4 min) timescales resulted, as expected, in relatively noisy SOFAST-HMQC spectra. Nonetheless they allowed the close monitoring of rapidly exchanging residues that occurred within the first few hours. Typical SOFAST-HMQC spectra at $t=0$ and $t= 4$ days (Fig.4.6.) illustrates the challenge that lies behind HDX-NMR studies of high molecular weight proteins such as α_1 -antitrypsin. Indeed, the relatively short number of scans early in the exchange resulted in line-broadening, and

initial peaklist. The 4 min SOFAST-HMQC experiments required for a fast monitoring of the exchange in intensity, unsurprisingly demonstrated a low sensitivity, as a consequence of both the large size of the protein, and experimental timescale. Even though remarkable given the size of α_1 -antitrypsin, the spectra low sensitivity rendered the transfer of the assignment across, a challenging process that was undertaken cautiously.

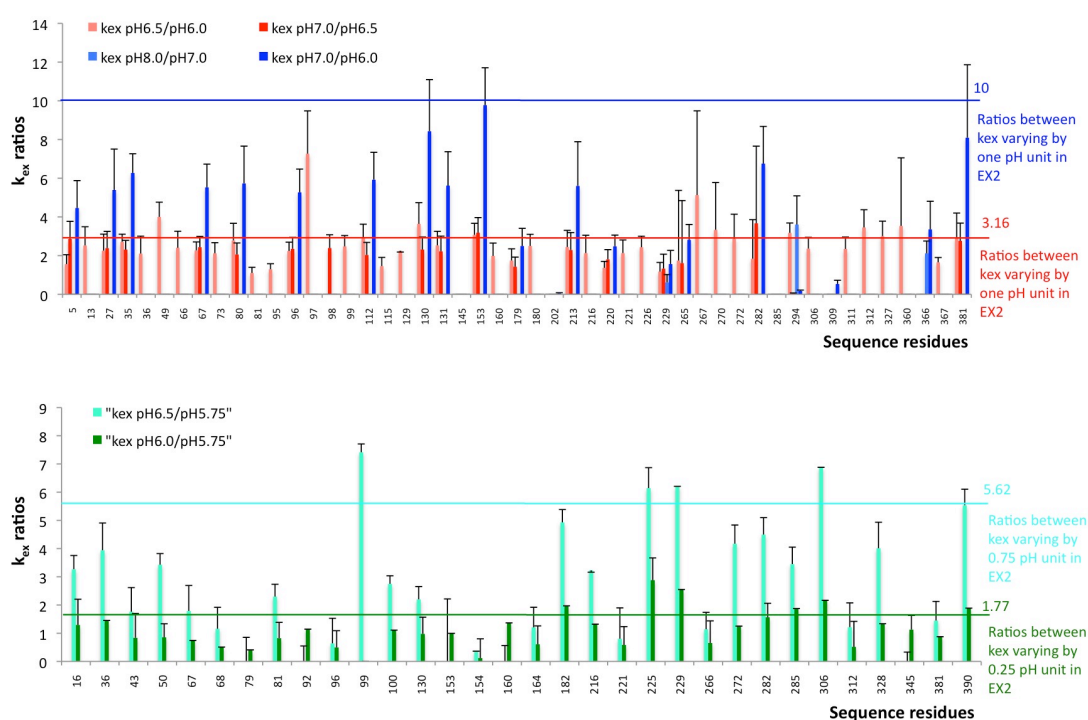


Figure 4.7. Characterisation of the EX2 regime. The k_{ex} ratios of pH 6.5/pH 6.0 (pink), pH 7.0/pH 6.5 (red), pH 8.0/pH 7.0 (light blue), pH 7.0/pH 6.0 (dark blue), pH 6.5/pH 5.75 (dark green), and pH 6.0/pH 5.75 were measured to assess whether residues reached EX2 to allow further protection factor calculation. Since $[H^+] = 10^{-pH}$, therefore ratios between one full pH unit, will be equal to 10, 0.75 pH units ratio corresponds to a value of 5.62, 0.5 pH units ratio corresponds to a value of 3.16, and 0.25 pH units ratio corresponds to a value of 1.77. Residues that reached the EX2 phase, by reaching the corresponding theoretical ratio, were further investigated for protection factors and free energy calculations.

Analysis of exchange curves using exponential fitting was possible for approximately 50 residues. When exchange is directly proportional to pH ('EX2 regime'), accurate protection factors can be calculated (Fig.4.7). The EX2 regime was assessed by quantifying the k_{ex} ratios between the various pH conditions

used in this analysis, and the subsequent theoretical values corresponding to linear dependency between the exchange rate and the pH. Since the exchange is directly proportionate to pH in EX2 regime, and since the difference in proton concentration between different pH conditions is equal to $10^{-\Delta\text{pH}}$; the theoretical ratios for a residue to be in the EX2 regime can be characterised as described in Fig.4.7. The exchange behaviour of residues identified as being in EX2 is described further below.

This study identified three categories of HD exchange occurring within native α_1 -antitrypsin: (i) Slow exchange (observed by constant crosspeak intensity across the timespan measured of 4 days, suggesting a highly structured environment and consequent protection from deuterated solvent, Fig.4.8. set1) (ii) crosspeak exchange observed as exponential intensity decay ($I_{(t)} = I_0 \cdot e^{(-k_{\text{ex}} \cdot t)}$; Fig.4.8. set2); and (iii) fast exchange (complete resonance exchange occurring during the assay deadtime, Fig.4.8. set3). For the residues entering an EX2 regime (see below for these), subsequent calculation of protection factors and ΔG values at a residue-specific level was undertaken.

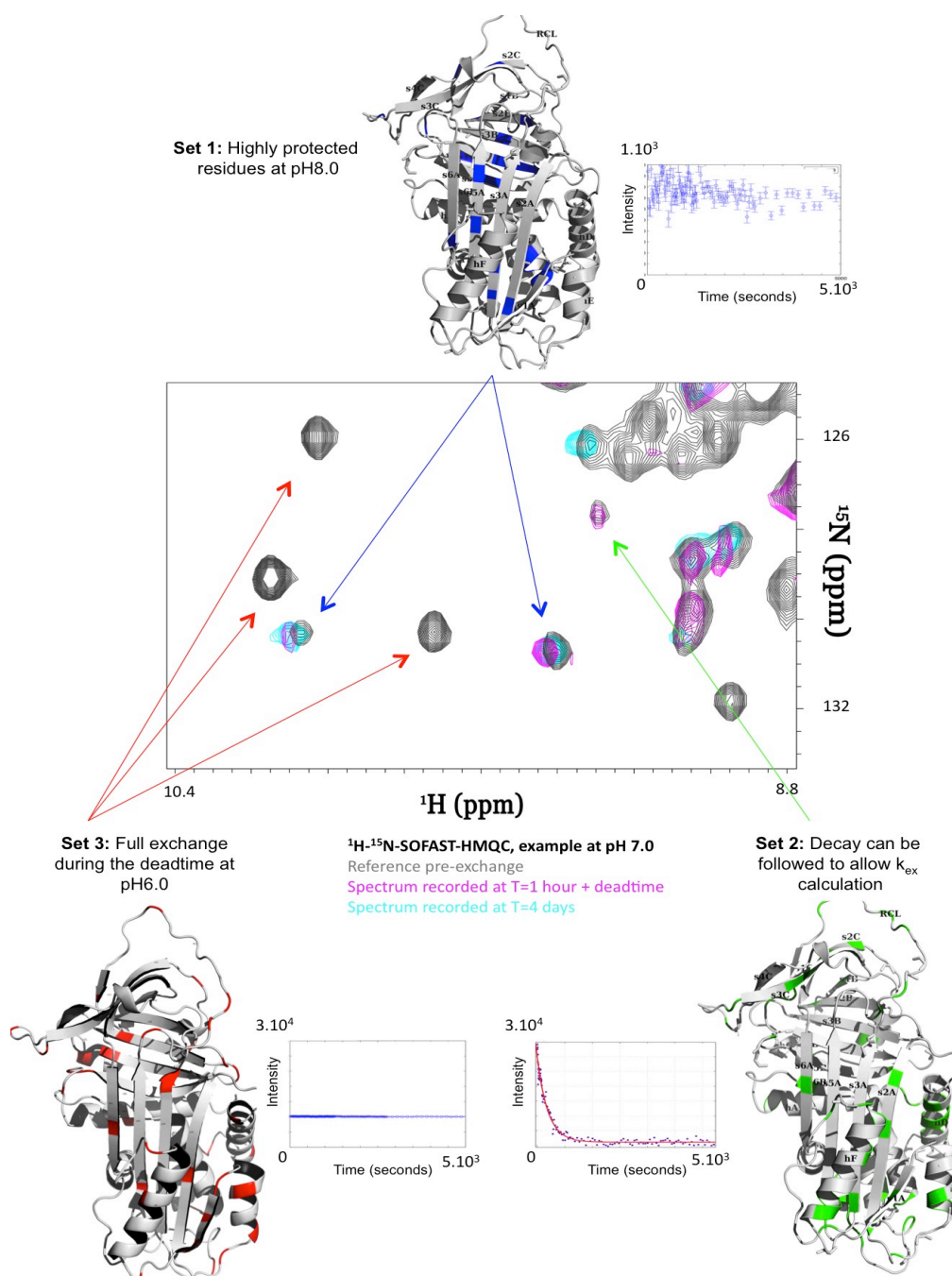


Figure 4.8. Defining residue protection in native α_1 -antitrypsin. Centre: Typical SOFAST-HMQC spectral overlay from HDX-NMR experiment at pH 7.0 over time ($t = 0$, in grey; $t = 1$ hour, in magenta; and $t = 4$ days, in cyan, SOFAST-HMQC). Contour levels standardised for the number of scans in each experiment. Top: Highly protected residues showed no intensity decay throughout the duration (4 days) of the experiment at the highest pH studied (pH 8.0, set1). Lower right: Residues showing higher lability show the loss of their own crosspeak intensity during the experimental deadtime (5-7 minutes) even at the lowest pH studied (pH 6.0, set 3). Lower left: The residues in set 2 are those for which the decay over time was fitted with an exponential function allowing determination of k_{ex} values. An example of a typical SOFAST-HMQC spectrum undergoing HDX-NMR exchange at pH 7.0 was illustrated in the centre part, demonstrating examples of these three different behaviours from a spin dynamic prospect. Residues belonging to each set are illustrated by representative 'decay' curves.

Approximately 50 crosspeaks showed k_{ex} changes consistent with an EX2 regime over the titration (Fig.4.9.). Under these conditions, the exchange rate k_{ex} probes for the thermodynamic equilibrium for the opening-closing of the structure (quantified by the protection factor P): $P = k_{cl}/k_{op} = k_{int}/k_{ex}$. P relates to the difference in free energy of the observed residue, $\Delta G_{app} = -RT\ln(k_{op})$ where $k_{op} = 1/P$.

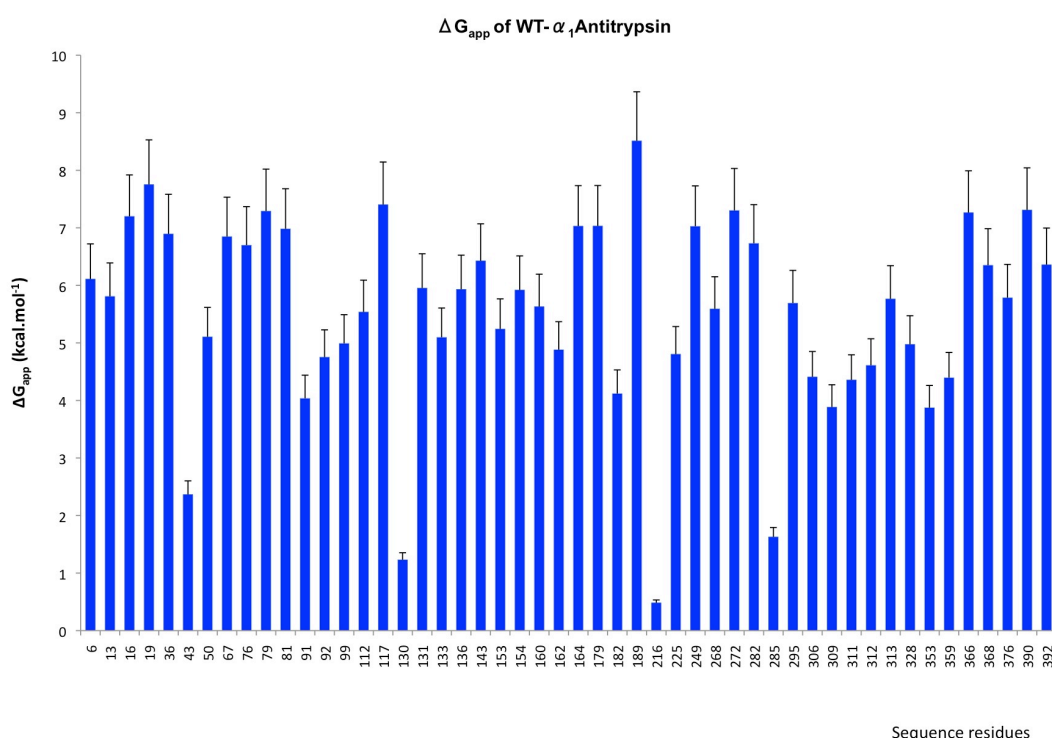


Figure 4.9. HDXNMR of α_1 -antitrypsin: ΔG_{app} as a function of the residues number HDXNMR experiments on α_1 -antitrypsin allowed 200 well dispersed, non-overlapped backbone amide signals to be followed across the pH range between pH 8.0 and pH 5.75. The data inform upon the stability of the corresponding residues in the solution native state. Out of these, approximately 50 residue demonstrated an exchange rate that followed a linear dependency with pH, consequently allowing calculation of free energies derived from their protection factor values using the relation $\Delta G_{app} = -RT\ln(k_{op})$, as plotted against the sequence of residues. The error in the rate of exchange were obtained by the ratio of errors calculated for the rate of exchange: $(a/x+b/y)x(a/b)$ (for errors $x \pm a/y \pm b$). The free energy errors were derived from the rate of exchange errors where $\theta_{kop} = \theta_{kex}/k_{ex}^2$ and $\theta_{\Delta Gapp} = \theta_{kop}/k_{op}$. (as was described in Chapter II, Section 2.2.4.4.).

4.3. Discussion

The HD exchange data derived in this study allowed us to assess dynamics of the native state of native wild-type α_1 -antitrypsin in solution in the millisecond to second timescale range. Globally, the lability (Fig.4.10.) ranged between $0.4 \text{ kcal.mol}^{-1} \geq \Delta G_{\text{app}} \leq 9 \text{ kcal.mol}^{-1}$. This is broadly consistent with global ΔG values derived for native wild-type α_1 -antitrypsin by other experimental methods⁹². This range should be approximately comparable to the 0-100% protection scale used to interpret the earlier HDXMS studies of native α_1 -antitrypsin in solution (Fig.4.10.A) ⁹².

Overall, our data indicate that residues demonstrate high rigidity or high lability on this scale, with relatively few residues reporting intermediate levels of protection, the differential findings are consistent with those obtained by HDXMS (Fig.4.10.B). Although the HDXNMR data covers fewer residues, the increased resolution indicates that exchange behaviour may be more polarised at the residue-specific level around the native conformer. Alternatively, a similar pattern could arise from thresholding effects.

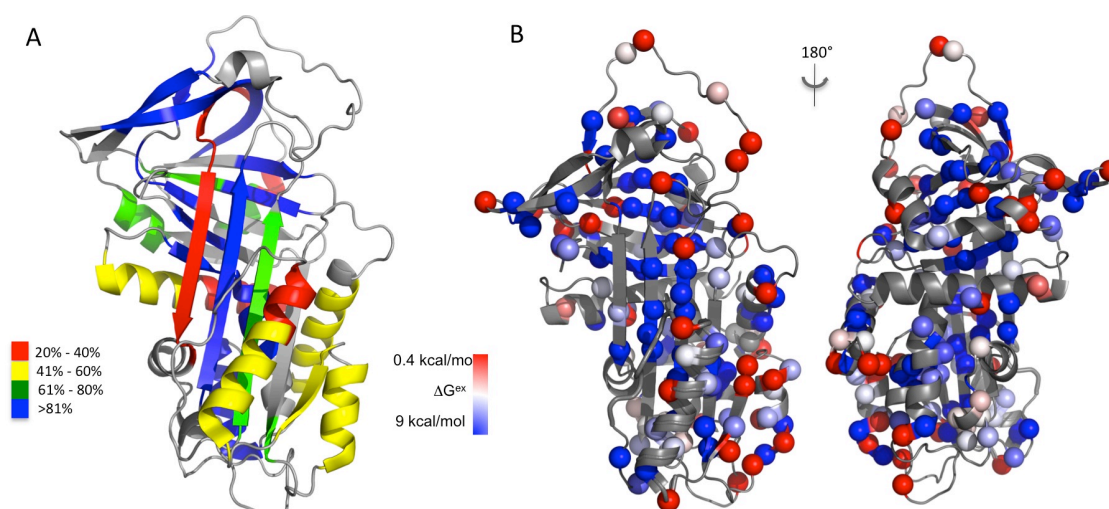


Figure 4.10. Stability indices derived by HD exchange observations from MS and NMR spectra A- HDXMS data plotted on the structure of native α_1 -antitrypsin (PDB ID 3NE4) with reported peptide regions coloured according to calculated degree of protection⁹²; B- HDXNMR data plotted on the same structure. Spheres identify the residues that were unambiguously assignable and for which ΔG values could be attributed relative to the quantifiable range in this HDXNMR analysis. $\Delta G \geq 9 \text{ kcal.mol}^{-1}$ are represented in blue (stable), and where $\Delta G \leq 0.4 \text{ kcal.mol}^{-1}$ are represented in red (labile).

More specifically, the data define a region of high HD exchange protection extending from β -sheet C through the core (β -sheet B and B-helix) through to the middle of β -sheet A. Breaking this down further (Fig.4.11.) the interactions suggest 2 hydrophobic clusters. Cluster 1 locks closed the centre of β -sheet A (strands 3 and 5, residues Phe119, Ala183, Val185, Tyr297, Lys335, Leu338) by interaction with the B-helix (Ala60, Phe61, Met63, Leu64) and the lower, buried strand residues in β -sheet B (Lys368, Leu373, Met374, Phe384). This corresponds to the shutter region, and boundaries of s4A/loop insertion site^{77; 140}. Cluster 2 affects hydrophobic residues interacting in a β -barrel region between β -sheets B and C. Destabilisation of this region by 'latch mutations' was recently described as a pathogenic mechanism in mild deficiency variants of α_1 -antitrypsin¹⁹¹. Latch interactions constrain the RCL by stabilising the closed conformation of the 'gate' (s4C-s3C), their loss may facilitate release of s1C, enabling further conformational change and intermediate formation¹⁹¹. This region is also proposed as a nucleation site for protein folding⁹⁰.

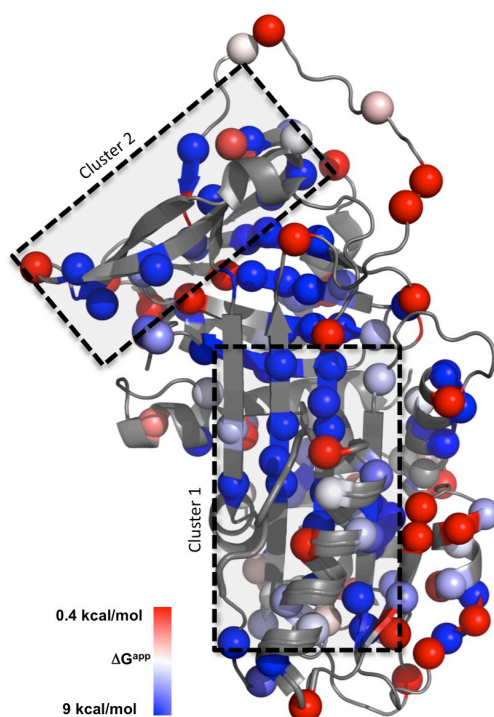


Figure 4.11: Stability Hydrophobic clusters highlighted by HD exchange observations for residues within the structure of native α_1 -antitrypsin (PDB ID 3NE4) with alpha helices and beta strands colored according to the ratio of the number of slow-exchanging hydrogens (experiment)/the number of protected hydrogens highlighting two hydrophobic clusters involving the β -sheet A and the β -sheet C.

The minimally-protected, flexible regions include the RCL and parts of the F-helix, which are known to interact with β -sheet A to open the s4A site. This suggests that the dominant regulation of opening is by cluster 1 interactions rather than conformational stability of the reactive loop or F-helix. The effects of disease mutations support this. Whilst mutations shown or predicted to destabilize these regions do cause polymerisation and circulating deficiency^{192; 193; 194} the effects of destabilizing s5A and shutter region mutations are much greater^{95; 173; 195}.

Overall, HDXMS of the native state of α_1 -antitrypsin demonstrated global metastability was related to a bimodal pattern of labile and protected residues around the molecule. Consequently, deficiency mutations in different positions in the molecule lead to different levels of severity of the disease. Those affecting residues that are inherently more labile (e.g. around and beyond the C-terminal of the RCL¹⁹¹, and the F-helix) have milder consequences than those that affect more protected residues (e.g. around breach⁸⁸ and shutter¹⁹⁶ regions).

Together these data provide the basis for further studies of dynamics at other timescales e.g. by relaxation studies that report upon ps-ns motions (see Appendix A for my introductory work in this area). They also represent the baseline readout against which to compare the effects of disease-relevant mutations upon the dynamics of monomeric α_1 -antitrypsin in solution over ms-s timescale ranges.

**Chapter 5 – CHARACTERISATION OF THE SOLUTION BEHAVIOUR OF
THE NOVEL MILD DEFICIENCY MUTANT Lys154Asn α_1 -ANTITRYPSIN**

5.1 Introduction

α_1 -antitrypsin is synthesized by hepatocytes and released in the plasma to protect the lung from neutrophil elastase. The most clinically significant mutation leading to α_1 -antitrypsin deficiency is the homozygote of Z-mutation (Glu342Lys), and is carried by 1 in 27 of North European population. Formation of polymers in Z- α_1 -antitrypsin deficiency is believed to occur via an intermediate conformation before polymers form and accumulate within the Endoplasmic Reticulum (ER) of hepatocytes, resulting in both a gain- and loss-of-function. As an inhibitory serpin, the function and dysfunction of α_1 -antitrypsin depends on the structures and dynamics of the conformers available to the polypeptide chain and the energy landscape that incorporates them. These are complex interrelationships. They have been tractable to systematic analysis by biochemical, biophysical and structural methods in large part because of insights derived from characterization of naturally occurring mutations associated with disease^{197; 198}. Such mutations typically disrupt regulation of the stressed (S, metastable) to relaxed (R, hyperstable) transition required for function. Thus the Z (Glu342Lys) mutation in α_1 -antitrypsin points up the importance of proximal hinge and upper s4A ("breach" region) behaviour⁸⁸. The cluster of polymerogenic mutations affecting interactions that stabilize closure of the middle of β -sheet A defines the regulatory importance of the so called "shutter region"¹⁹⁶. Other mutations point to a minor role for stabilization of the residues around and beyond the C-terminal of the RCL¹⁹¹. The discovery of a novel mutation affecting the F-helix of α_1 -antitrypsin and associated with mild deficiency but not overt disease (i.e. a *forme fruste* mutation) provided an opportunity to extend this principle to NMR studies of the conformational behaviour of α_1 -antitrypsin.

5.1.1. Case report

A 34 year-old man participated in family screening for α_1 -antitrypsin deficiency following the diagnosis of his daughter as a PiZZ homozygote. Initial iso-electric focusing (IEF) phenotyping of his circulating α_1 -antitrypsin indicated

he was also a PiZZ homozygote. However his circulating levels of α_1 -antitrypsin were only moderately reduced (median 0.6 mg/ml, normal range 1.5-3.5), i.e. higher than expected for a PiZZ homozygote (typical levels \sim 0.2 mg/ml). He was therefore genotyped at Queen's Medical Centre, Nottingham, UK (Prof N. Kalsheker and Dr M. Hill). This revealed him to be compound heterozygous for the Z (Glu342Lys) allele and a novel variant (Lys154Asn) of α_1 -antitrypsin. This was named, the Queen's variant of α_1 -antitrypsin, after the centre where the mutation was identified.

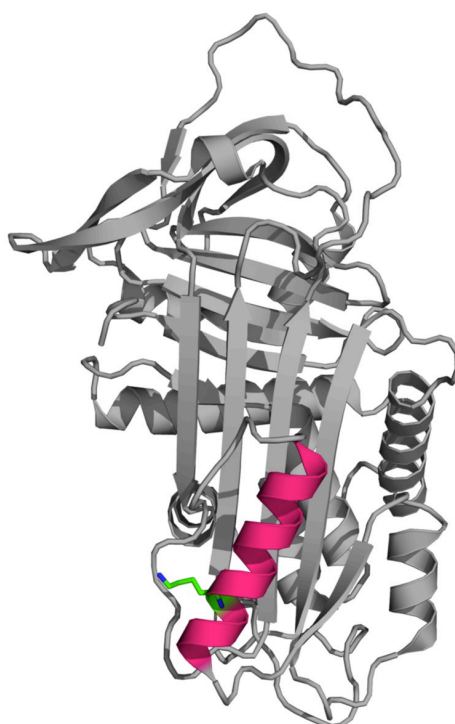


Figure 5.1. Lys154Asn α_1 -antitrypsin mutation associated with mild deficiency
The Lys154Asn (green) mutation is located on the F-helix (magenta) and highlighted here on the crystal structure of wild-type native α_1 -antitrypsin (PDB:3NE4)⁸⁸.

5.1.2 Involvement of the F-helix in conformational transitions

The Lys154Asn mutation arises within the F-helix (hF) of α_1 -antitrypsin (Fig. 5.1). The F-helix lies across β -sheet A and with which it shares a number of interactions⁸⁸. Its N-terminus arises from the base of strand 1 of β -sheet A (s1A) and it is comprised of four turns. The C-terminal precedes a linker region that runs antiparallel to hF before linking to the base of s3A. The helix and linker

share a number of polar interactions and together enclose a hydrophobic interface with the underlying part of β -sheet A that is further stabilized by a number of polar interactions⁸⁸. Several lines of evidence indicate this motif remodels during serpin conformational transitions and that it may play a role in regulating the opening of β -sheet A.

5.1.2.1 Rouen-VI antithrombin

Firstly, almost 2 decades ago, Perry and colleagues characterized the Asn187Asp mutant of antithrombin (Rouen-VI variant)^{199; 200}. The mutation disrupts an hF-linker interaction that is conserved within the serpin superfamily and allows the spontaneous conversion of native antithrombin to inactive latent and polymerised conformations. These data support a role for the interaction in holding β -sheet A in its closed, native conformation.

5.1.2.2. Leu55Pro α_1 -antichymotrypsin

The Leu55Pro mutation in the serpin α_1 -antichymotrypsin that is closely related to α_1 -antitrypsin is associated with circulating deficiency and emphysema. Biochemically, the mutant behaves characteristically for a shutter region variant, favouring polymerisation and adoption of the latent conformation. However it also facilitates formation of the alternative monomeric δ conformation. X-ray crystallography demonstrates how this conformation has an expanded β -sheet A with the s4A site filled by remodeling of residues from 2 distinct motifs (Fig. 5.2.; PDB: 1QMN)⁷⁷.

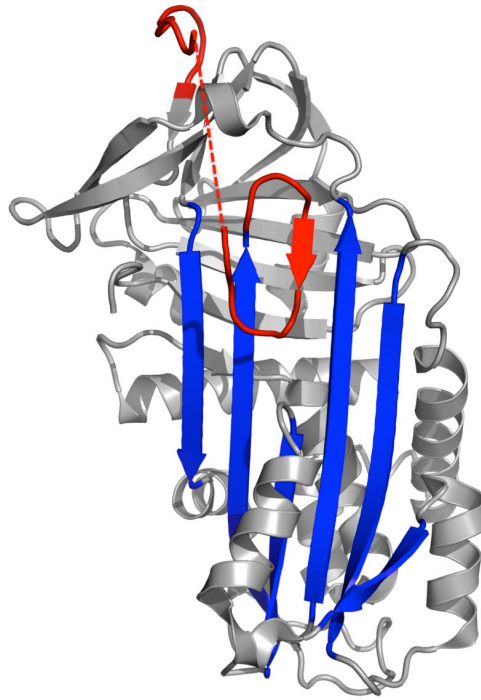


Figure 5.2. α_1 -antichymotrypsin serpin in the delta conformation (partial loop insertion). Inactive conformation of the serpin α_1 -antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease. (PDB: 1QMN). The dashed line represents the parts that aren't seen in the crystal structure.

The upper s4A (breach) site is occupied by partial insertion of the reactive centre loop. Fascinatingly, however, the lower s4A site is filled by unwinding of the last turn of F-helix to allow the linker residues Asn-163 to Thr-170 to adopt the requisite β -sheet structure⁷⁷. Leu55Pro α_1 -antichymotrypsin adopts the δ conformation in parallel with latent and polymeric states. The δ conformation has been interpreted as a trapped form of the common intermediate (M^*) along a branched pathway from which all three conformations raising the possibility that hF remodeling was involved in intermediate formation.

5.1.2.3. Solution studies of F-helix remodeling

The prospect of hF remodeling in conformational transitions is also implied by the structure of the serpin-enzyme inhibitory complex⁶⁵. In the absence of such remodeling the F-helix represents a physical impediment to the rapid insertion of the RCL and translocation of the covalently bound target protease that is required for inhibitory function²⁰¹. This hypothesis was therefore explored by Cabrita *et al.* who used alanine and tryptophan mutagenesis studies to probe F-helix stability and conformational change of this helix during proteinase inhibition, serpin folding, and polymerisation^{192 193}. Taken together, the data from these studies demonstrated that F-helix undergoes different degrees of remodeling during these different conformational transitions. A reversible conformational change was observed in both N- and C-termini of hF during inhibition pathways, but only the C-terminus reported changes during polymerisation. Conversely, unfolding was associated with changes throughout the length of the helix. These studies were further complemented by HD exchange mass spectrometry data highlighting lability of the upper hF region in the native state of α_1 -antitrypsin⁹².

The characterization and X-ray crystal structure of Gly117Phe α_1 -antitrypsin also promoted the importance of F-helix as a motif that could propagate conformational change between upper and lower s4A opening¹⁷⁰. Conversion from the native conformation to the M* model derived from the structure of δ Leu55Pro α_1 -antichymotrypsin entails potential steric clashes with the upper hF-linker motif as configured in the native state. These clashes can be resolved by transient remodeling of the hF-linker region prior to opening of the lower s4A site. Consistent with this hypothesis, native Gly117Phe α_1 -antitrypsin is dramatically stabilized against polymerisation in association with a half-turn downward shift in hF. The shift will predictably prevent the induction of such remodeling and therefore reduce the tendency of the lower s4A site to open.

5.1.2.4. Potential structural effects of the Lys154Asn mutation in α_1 -antitrypsin

Taken together, these data strongly support a role for the hF+linker motif mediating functional and pathological conformational transitions in α_1 -antitrypsin and other serpins. Lys154Asn α_1 -antitrypsin was associated with mild contribution to circulating deficiency in a compound heterozygote and no clinical disease. These considerations led to the hypothesis that the mutation would destabilize the hF+linker motif in α_1 -antitrypsin to give a mildly polymerogenic biochemical phenotype. Analysis of the high-resolution crystal structures of the native state of α_1 -antitrypsin indicated that such an effect could arise by destabilizing one or more of 3 polar interactions (Fig.5.3.). These were probed in a mammalian cell model and the biochemical phenotype of the Queen's variant characterized following recombinant expression in *E. coli*.

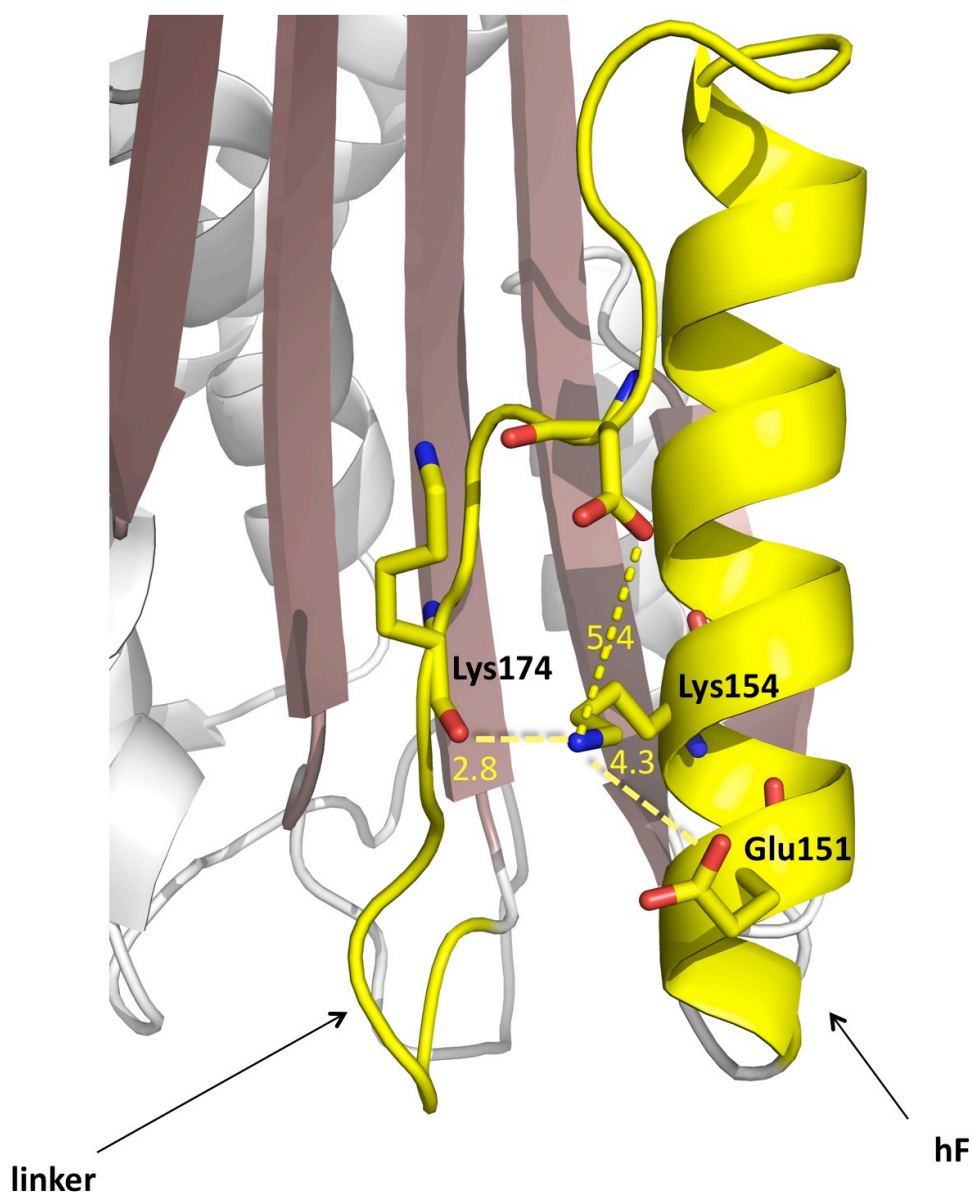


Figure 5.3. Potential structural effects of the Lys154Asn mutation in α_1 -antitrypsin. The Lys154 side chain participates in 2 polar interactions, a salt bridge to Glu151 nearby on hF and a hydrogen bond to the main chain carbonyl of Lys174.

5.2. Results

5.2.1. Production of recombinant Lys154Asn α_1 -antitrypsin

Lys154Asn α_1 -antitrypsin was expressed recombinantly and purified as described in Chapter II (Figure 5.3). Monomeric Lys154Asn α_1 -antitrypsin eluted

intact and without contaminant from the second, Q-sepharose column in a single peak.

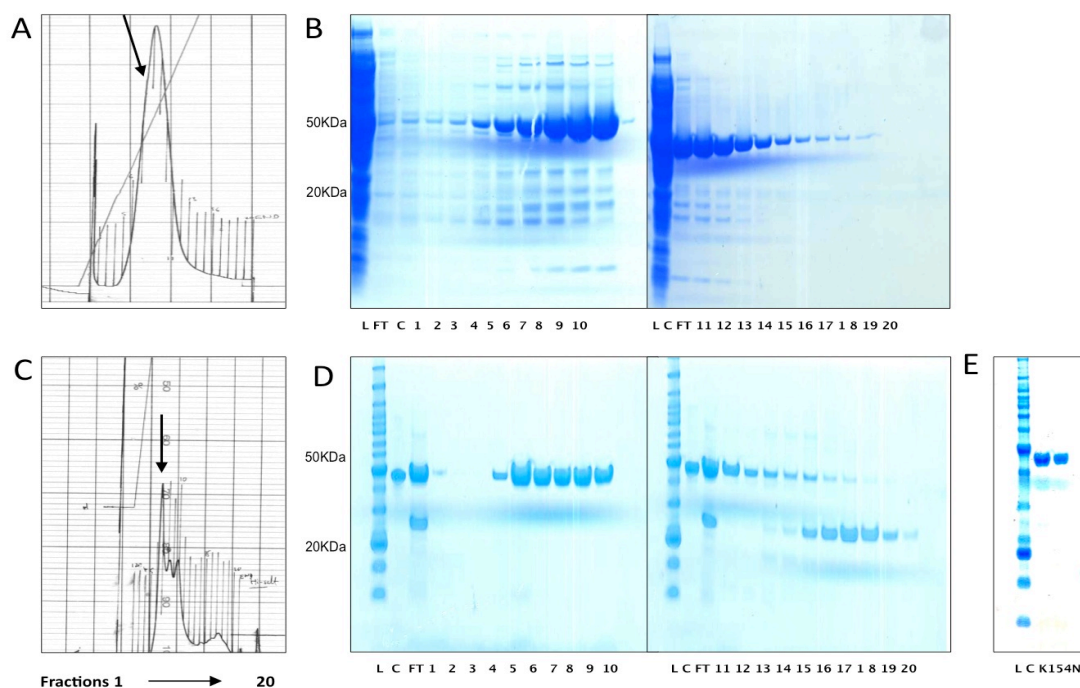


Figure 5.3. Lys154Asn α_1 -antitrypsin Ni-sepharose - Q-sepharose purification. Gel labels: L: ladder of molecular weight markers; C: Control wild-type α_1 -antitrypsin; numbers refer to fraction order; FT: Flow-through of material not binding to the column during loading, 10 μ l of fractions and flow-through loaded per lane). (a) Elution profile from initial high affinity nickel-sepharose column, the peak representing the elution profile of Lys154Asn α_1 -antitrypsin is shown by the arrow; (b) composition of the 20 fractions obtained after nickel-sepharose purification shown by SDS-PAGE electrophoresis Typically monomeric Lys154Asn- α_1 -antitrypsin eluted at 50-150mM imidazole. (c) High-affinity Q-sepharose elution profile; (d) composition of fractions obtained after Q-sepharose purification assessed by SDS-PAGE electrophoresis otherwise loaded and labelled as in (b). (e) SDS-PAGE of final Lys154Asn α_1 -antitrypsin product.

5.2.2. Biochemical characterization

Conformational homogeneity of the Lys154Asn α_1 -antitrypsin product was assessed by 7.5% (v/v) non-denaturing PAGE (Fig.5.4.). Lys154Asn α_1 -antitrypsin demonstrated the doublet appearance characteristic of native α_1 -antitrypsin purified from this system. A trace of a slower migrating species was

also apparent (shown by *), consistent with a dimer or monomeric intermediate conformation. This band varied in intensity (both waning and waxing) between successive gels from the same aliquot, and was occasionally apparent in the wild-type material also. Native MS experiments (detailed below) also indicated there was no significant population of dimeric material in either wild type or Lys154Asn α_1 -antitrypsin products. Taken together these data suggested the extra band represented a monomeric intermediate state observable upon native electrophoresis, whose stability relative to the native state was very sensitive to electrophoresis conditions.



Figure 5.4. 7.5% (v/v) non denaturing PAGE of Lys154Asn α_1 -antitrypsin. Both wild-type (WT) and Lys154Asn (K154N) α_1 -antitrypsin products predominantly migrate consistent with recombinant protein in the native conformation. A trace band of material with a slower migration pattern was also observed (*) for the Lys154Asn mutant (and occasionally for the wild-type) that varied in intensity between experiments without a temporal trend.

TUG-PAGE unfolding of native Lys154Asn α_1 -antitrypsin confirmed its metastability (Fig.5.5.) and indicated it was more conformationally heterogeneous than the wild-type protein under low urea concentration electrophoresis conditions. Once again an intermediate state was evident under these conditions (indicated by *). However the profile for Lys154Asn α_1 -antitrypsin showed less cooperative unfolding than the wildtype protein,

consistent with prolonged population of an intermediate species. Furthermore nondenaturing and TUG-PAGE revealed that Lys154Asn α_1 -antitrypsin populated a state that migrated more slowly than the native conformer. It persisted in reducing conditions but not following SDS denaturation, and it was only minimally populated by the wildtype protein.

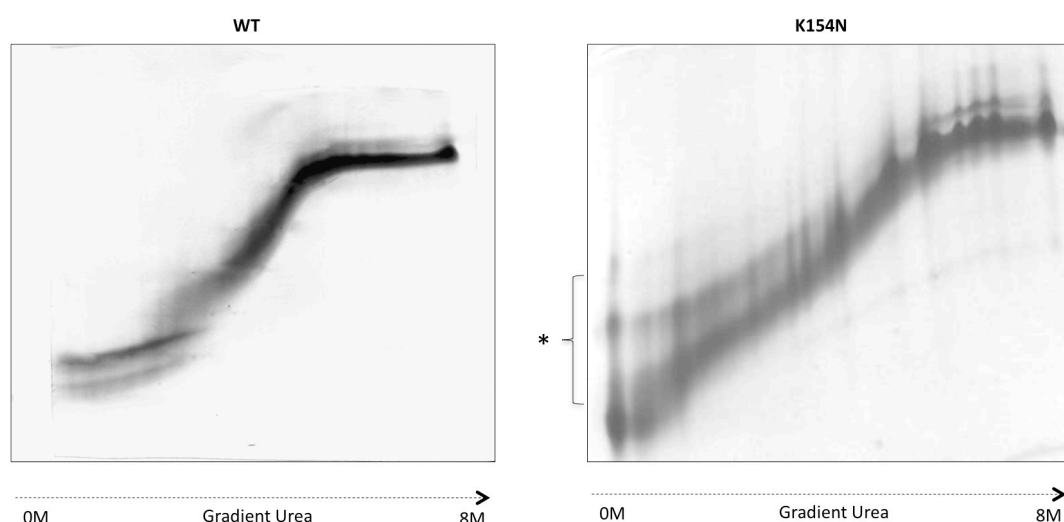


Figure 5.5. Transverse Urea Gradient (TUG) electrophoresis of wild-type and Lys154Asn α_1 -antitrypsin. TUG gels loaded with 50 μ g protein shows the typical metastable behaviour for the wild-type (WT) sample, while the Lys154Asn variant indicates population of an intermediate ensemble at low concentration of urea as well as some formation of latent and oligomeric material.

5.2.3 Activity assay and inhibitory kinetics

The Lys154Asn mutant had 67 % of the functional activity (Fig.5.6.) of wildtype α_1 -antitrypsin (i.e. 1.6-fold greater stoichiometry of inhibition (SI)). Its apparent association rate constant (k_{app}) with bovine α -chymotrypsin was $3.3 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$, 47 % of the value for the wild-type protein. When corrected for the increase in SI, the association rate constant (k_{app}) for Lys154Asn α_1 -antitrypsin was $7.3 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$ ($n=3$), 63 % of the value of the wild-type control ($1.16 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$)¹⁹³.

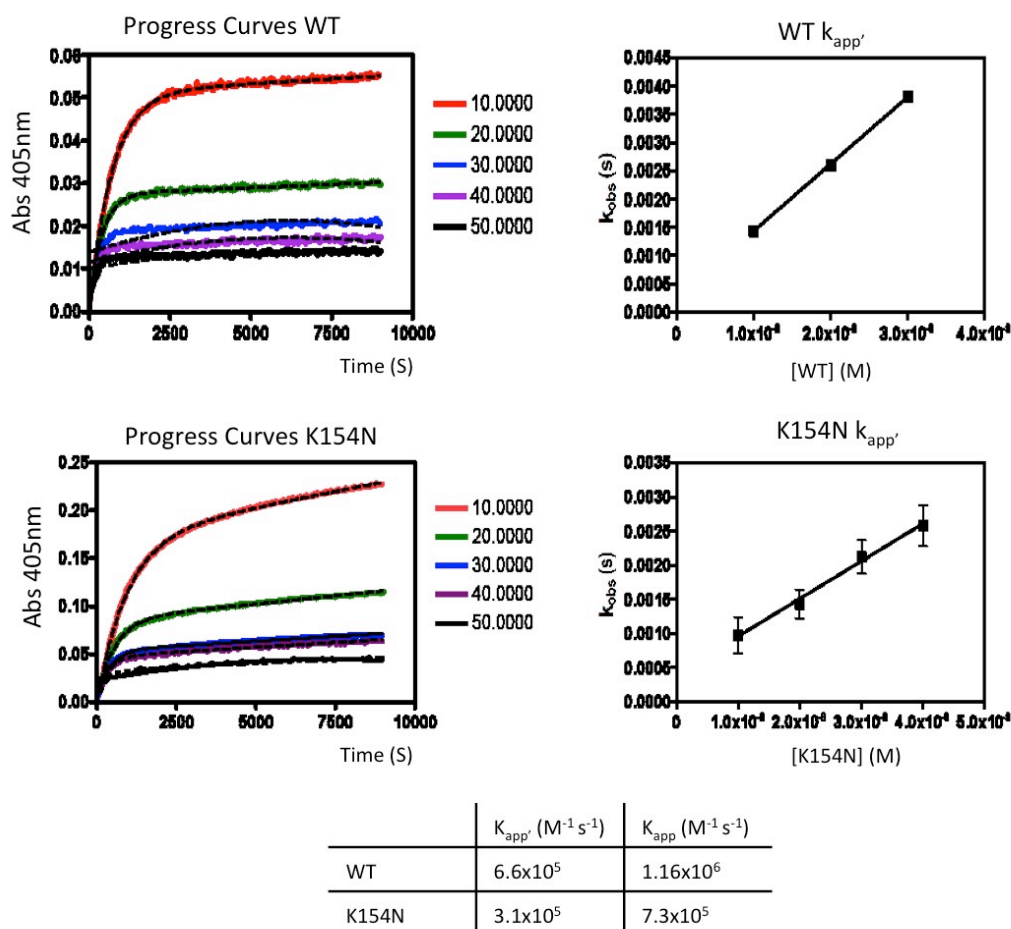


Figure 5.6. Kinetics of α_1 -antitrypsin. The curves (left) are the colourimetric readouts of substrate breakdown by chymotrypsin over time, in the presence of increasing α_1 -antitrypsin indicated by line colour key. The curve fit data can be processed to give observed association rate constants (k_{obs} , in seconds) values for α_1 -antitrypsin with chymotrypsin specific for each α_1 -antitrypsin concentration (relationship plotted on the right, $n=3$). The slope of the line gives the general apparent association rate constant (k_{app}') that holds true across all α_1 -antitrypsin concentrations (in $M^{-1}.s^{-1}$).

SDS-PAGE and enzyme inhibitory activity assay, confirmed increased substrate-like behaviour in Lys154Asn α_1 -antitrypsin relative to the wild-type, characterised by concomitant loss of activity (figure 5.7).

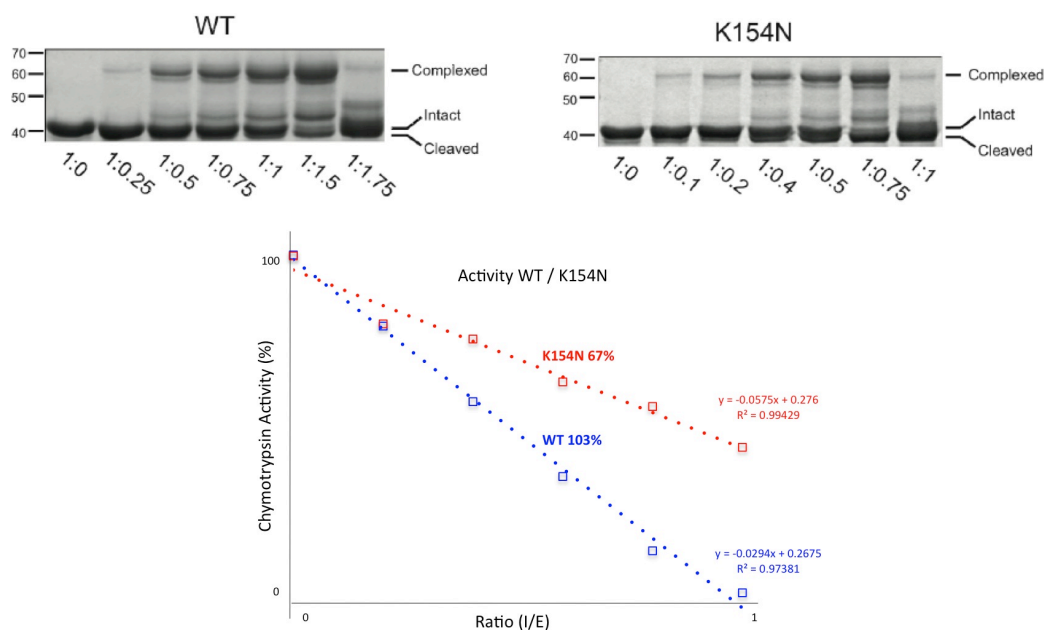


Figure 5.7. Enzyme inhibitory activity of the wild-type compared to the Lys154Asn disease mutant.

(A) 10% SDS-PAGE of wild-type (WT, left) and Lys154Asn (Lys154Asn, right) α_1 -antitrypsin interactions with bovine α -chymotrypsin. Molecular weights (kDa) are indicated by marker bars. Molar ratios of α_1 -antitrypsin:chymotrypsin are shown beneath each lane. (B) Inhibitory activity of the wild-type compared to the disease mutant Lys154Asn by assessing the chymotrypsin activity in (%) against the Ratio (Inhibitor/Enzyme), demonstrating an activity of 67% for the Lys154Asn relative to the wild-type.

5.2.4. Thermal stability and polymerisation

[work carried out in collaboration with Lakshmi Segu (Gooptu group, ISMB, UCL/Birkbeck, University of London) and Dr Benoit Roussel (Lomas group, CIMR, University of Cambridge)]

5.2.4.1. Circular Dichroism

To investigate changes in the thermal denaturation transition of α_1 -antitrypsin induced by the Lys154Asn mutation I undertook CD spectroscopy studies of the wild-type and variant protein (Fig.5.8.). These indicated that in the mutant the fold underwent a biphasic transition, with an initial subtle change between 35 °C and 45 °C before a more dramatic transition between 50 °C and 60 °C. Increasing concentration did not affect the first transition but caused the second transition to commence at a lower temperature. Wild-type α_1 -antitrypsin remained stable to 60 °C and underwent a sharp unfolding transition at higher

temperatures, consistent with previous studies²⁰². Increasing concentration cold-shifted the second half of the transition but did not affect its initial phase. Increasing protein concentration 2.5-fold (from 0.4 to 1.0 mg/ml) reduced the midpoint of the overall transition for wild-type α_1 -antitrypsin by 1.5 °C and that of Lys154Asn α_1 -antitrypsin by 1.0 °C. The profiles of the first part of the transition observed for wild-type α_1 -antitrypsin and the first transition observed for Lys154Asn α_1 -antitrypsin are both concentration-independent. Therefore, this likely reports upon structural change within the molecule. The subsequent transition is concentration-dependent. It therefore likely reports, to some degree, upon polymerisation of a species populated during the first phase since polymer formation occurs via population of a monomeric intermediate state that is polymerogenic¹⁰³.

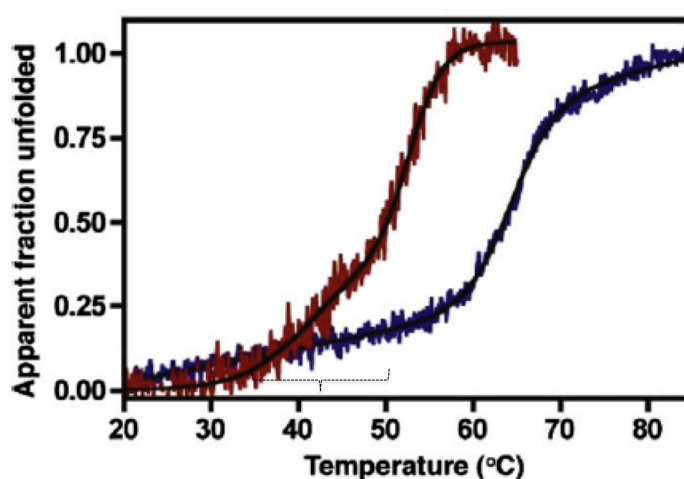


Figure 5.8. Biphasic thermal transition of Lys154Asn α_1 -antitrypsin observed by CD spectroscopy. Thermal denaturation CD spectroscopy (mean ellipticity at 222 nm, $n = 10$) for wild-type (blue) and Lys154Asn (red) α_1 -antitrypsin. Wild-type α_1 -antitrypsin starts unfolding at ~60 °C, while the Lys154Asn α_1 -antitrypsin mutant unfolds in a biphasic manner beginning at ~30 °C. The dashed-bracket highlights the initial phase (~35-50 °C, concentration independent) consistent with population of an intermediate ensemble in conditions including the physiological-pyrexial temperature range (36.5–37.5 °C).

Native PAGE analysis of wild-type and Lys154Asn α_1 -antitrypsin incubated for 30 minutes over a similar temperature range to that observed by CD spectroscopy indicated that the variant was more thermally labile than the

wild-type (Fig.5.9.). Polymerisation was observed at ≥ 45 °C for Lys154Asn α_1 -antitrypsin compared with ≥ 60 °C for the wildtype protein (both at 0.5 mg/ml) in these conditions.

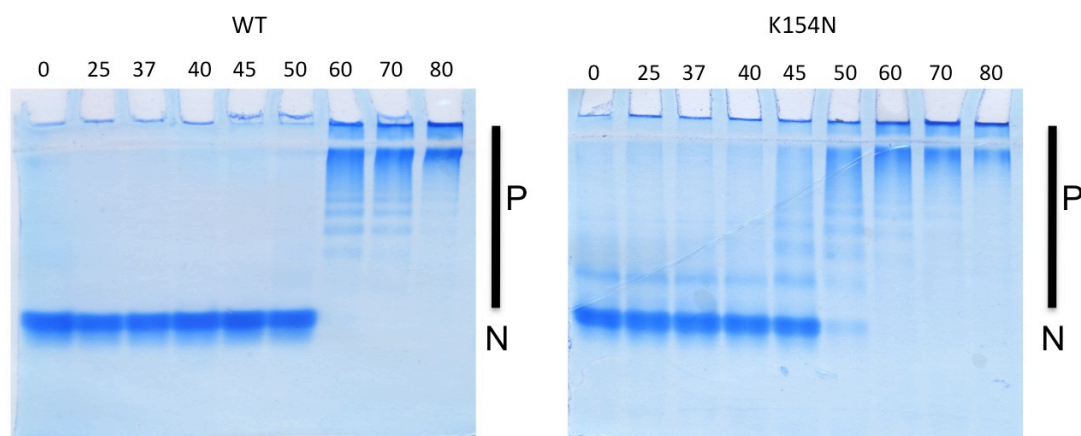


Figure 5.9. Thermal stability and polymerisation of α_1 -antitrypsin assessed by Native-PAGE. Thermal stability was assessed at 0, 25, 37, 40, 45, 50, 60, 70, 80°C by incubating WT and Lys154Asn α_1 -antitrypsin samples for 2 hours prior to assessing the thermal stability and polymerisation using non-denaturing Native-PAGE.

Polymer formation by Lys154Asn α_1 -antitrypsin between room temperature and 50 °C was next monitored by 2C1 monoclonal antibody ELISA (Fig.5.10.) and compared with results for wild-type α_1 -antitrypsin control. Differences between wild-type and mutant were significant ($p < 0.05$ to $p < 0.005$) for all temperatures. The increase in signal relative to starting material reached significance for Lys154Asn α_1 -antitrypsin for incubations at 47.5 ($p < 0.05$) and 50 ($p < 0.005$) °C.

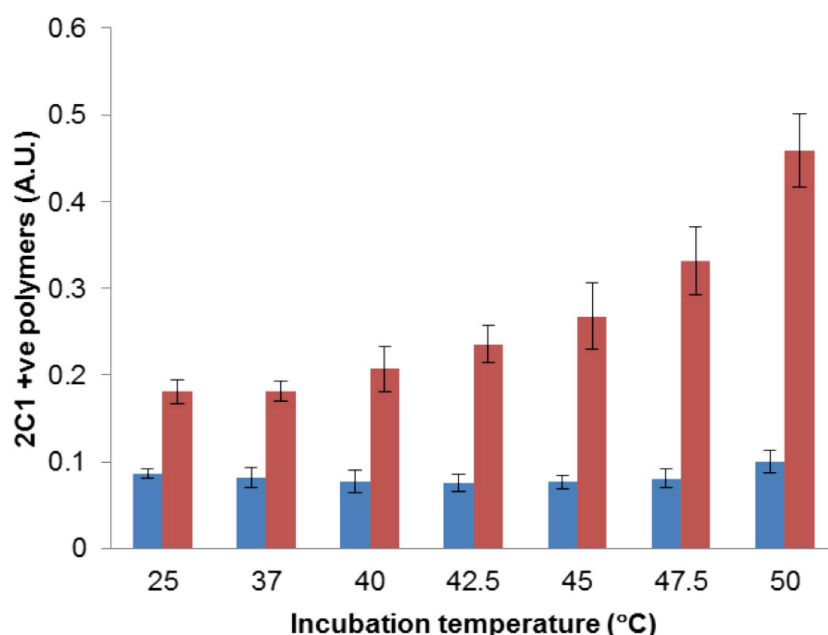


Figure 5.10. Formation of polymers recognised by 2C1 mAb by Lys154Asn α_1 -antitrypsin. Polymers formed by 30 min incubation between 40 °C and 50 °C of wild-type (blue) and Lys154Asn (red) α_1 -antitrypsin quantified by 2C1 monoclonal antibody ELISA (calibrated using a standard curve of Z α_1 -antitrypsin polymer)¹⁸⁹.

5.2.4.2. Native-PAGE of polymerisation at 37 °C and 42 °C

I next wished to assess polymerisation of Lys154Asn α_1 -antitrypsin over time in the physiological-pyrexial (36.5–37.5 °C) temperature range (pH and α_1 -antitrypsin concentration also mimicking those found *in vivo*). Wild-type and Lys154Asn α_1 -antitrypsin was therefore incubated at 0.5 mg/ml pH 7.4 and either 37 or 42 °C for up to 12 days (Fig.5.11). Wildtype α_1 -antitrypsin was highly stable in monomeric form at incubation for 12 days at 37 or 42 °C. In contrast most Lys154Asn α_1 -antitrypsin polymerised after 6 days at 37 °C and within 1 day at 42 °C, confirming population of a polymerogenic intermediate in solution across this temperature range.

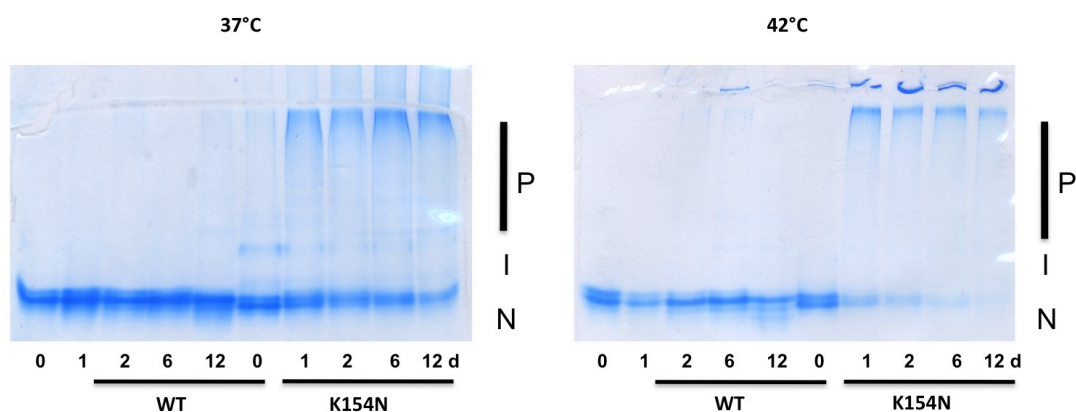


Figure 5.12. Polymerisation of α_1 -antitrypsin. Characterization of polymerization. 7.5 % non-denaturing PAGE analysis of time-dependent polymerisation at 37 °C and 42 °C for wild-type (WT) and Lys154Asn (K154N) α_1 -antitrypsin. N: native, I: putative intermediate, P: polymers.

5.2.4.3. Polymerisation in mammalian cells

pcDNA3.1 plasmids containing cDNA encoding wild-type or Lys154Asn α_1 -antitrypsin were transiently transfected into COS-7 cells. The resulting protein expression was characterized by western blot analyses of SDS- and nondenaturing PAGE. The polymer load was quantified by ELISA using the polymer-specific 2C1 monoclonal antibody specific to polymers observed in disease¹⁸⁹ as the primary antibody. The expression of M α_1 -antitrypsin resulted in very low levels of polymers intracellularly, while expression of the Z variant generated very high levels. Expression of Lys154Asn α_1 -antitrypsin resulted in far less polymerisation than seen with the Z mutant. (Fig.5.12)

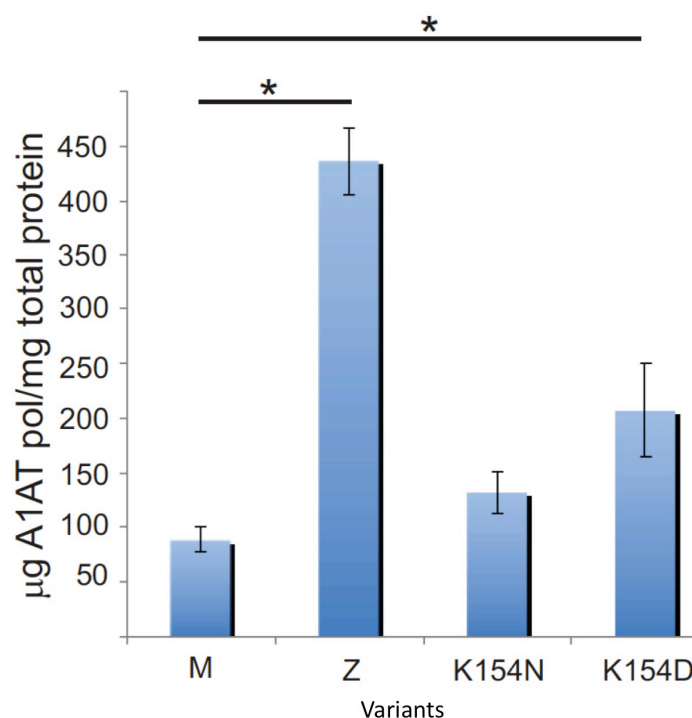


Figure 5.12. Biochemical characterization of polymerisation in the M, Z, Lys154Asn, Lys154Asp mutant α_1 -antitrypsin as reported by 2C1 monoclonal antibody recognition. Intracellular polymerisation of α_1 -antitrypsin following transient transfection of COS-7 cells. The polymer load was detected after 48 h of expression by ELISA using the 2C1 mAb. The data are the mean and SD of 3 repeats. Asterisks indicate differences with significance $p < 0.05$ (t test); n.s. indicates $p > 0.05$.

5.2.4.4. Polymerogenic intermediate formation assessed by ANS fluorescence

Polymerogenic intermediate formation in α_1 -antitrypsin has been previously studied by ANS fluorescence. ANS fluoresces strongly upon binding to regions in which both polar and hydrophobic motifs are exposed to solvent and so binds to intermediate states, rather than fully folded or unfolded polypeptides. The ability of the Queen's variant of α_1 -antitrypsin to bind ANS was therefore assessed (Fig.5.13.). Lys154Asn α_1 -antitrypsin showed hyperfluorescence compared with the wild-type protein at 25 °C. This was unexpected, since assays directly reporting structural information (CD spectroscopy, intrinsic fluorescence, and NMR spectroscopy) strongly indicated that both proteins were similarly well-folded and stable at this temperature. These observations of Lys154Asn α_1 -antitrypsin behavior are similar to those

described for Z α_1 -antitrypsin. Moreover, incubating Lys154Asn α_1 -antitrypsin with ANS for longer resulted in even greater hyperfluorescence, whereas increasing the temperature to 37 °C did not enhance this further. Wild-type α_1 -antitrypsin showed no change in ANS fluorescence at either temperature or upon prolonged incubation. These data indicate that for α_1 -antitrypsin variants in which conversion to the polymerogenic intermediate in solution is facile, ANS binding stabilizes the intermediate state sufficiently to skew the position of the native-intermediate equilibrium to the right. For Lys154Asn α_1 -antitrypsin, this effect is so significant that it outweighs any effect of temperature change between 25 °C and 37 °C. This supports the general view that the degree to which assays perturb solution equilibria must be considered when studying intermediate formation.

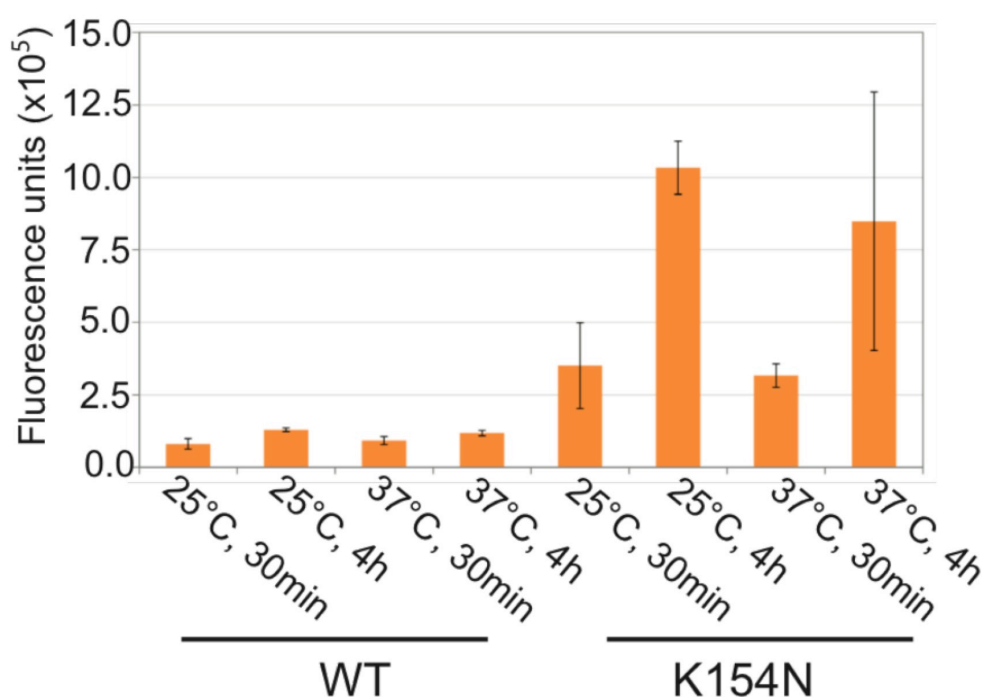


Figure 5.13. Lys154Asn α_1 -antitrypsin biophysical characterization, using ANS binding assay. Effects of mutation, temperature and incubation time of wild-type (WT) and Lys154Asn (Lys154Asn) α_1 -antitrypsin with ANS upon fluorescence at 480 nm wavelength (n=3 for each condition, error bars represent standard deviation values).

5.2.4.5. Ion Mobility-Mass Spectrometry

It was hypothesized that ion-mobility mass spectrometry (IM-MS) might have sufficient sensitivity to identify population of the intermediate while minimally perturbing the solution equilibrium. The measurement of the collision cross-section (CCS) monitored by ion-mobility mass spectrometry can be obtained by comparison of arrival time distributions, which indicates a change in conformation. It therefore provides an accurate measurement of conformational compactness. There has been a significant increase in the use of ion mobility mass spectrometry (IM-MS) to investigate conformations of proteins and protein complexes following electrospray ionization. Investigations which employ traveling wave ion mobility mass spectrometry (TW IM-MS) instrumentation rely on the use of calibrants to convert the arrival times of ions to collision cross sections (CCS) providing "hard numbers" of use to structural biology. Using this approach we measured the CCS at 20, 34 and 39 °C . At both 20 °C and 34 °C, monomeric Lys154Asn α_1 -antitrypsin was indistinguishable from the wild-type protein by IM-MS. However, at 39 °C, there was a 7.8 % increase in the CCS of monomeric Lys154Asn α_1 -antitrypsin relative to the wild-type protein in keeping with population of an intermediate state (Fig.5.14.). The CCS of the intermediate state has been calculated as ~18% greater than the native state⁹³ so a 7.8 % increase in CCS is consistent with substantial population (~ 40 %) of the intermediate state in equilibrium at physiological temperatures.

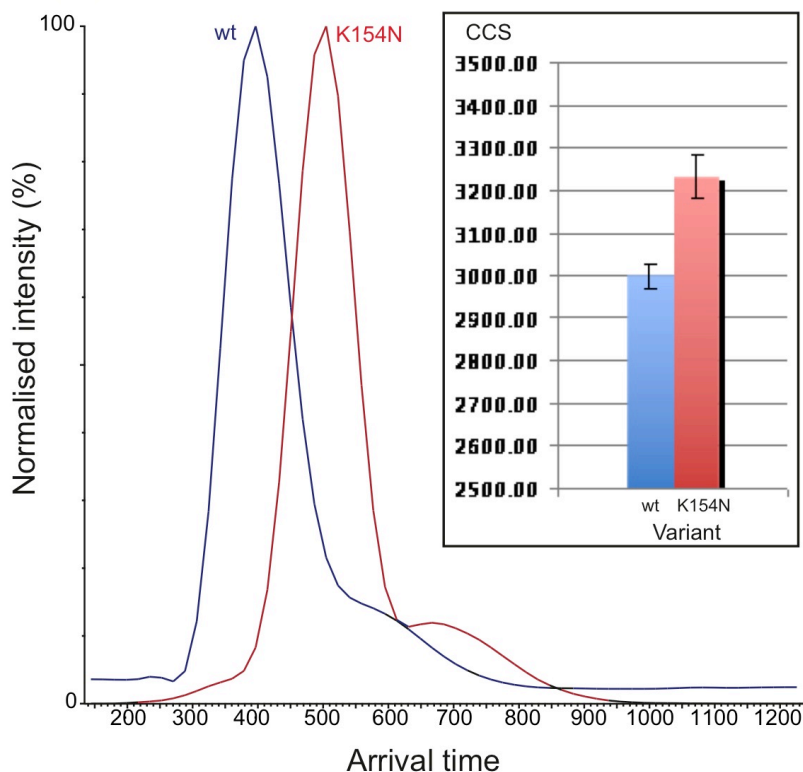


Figure 5.14. Lys154Asn α_1 -antitrypsin biophysical characterization, using Collision Cross Section studies (CCS) Arrival times and collision cross-section (CCS) values calculated by IM-MS for wild-type (blue) and Lys154Asn α_1 -antitrypsin indicated a 7.8 % increase in CCS in Lys154Asn relative to wild-type α_1 -antitrypsin at 39 °C (no difference was seen at 20 °C and 34 °C). Data are mean \pm standard deviation (error bars) of three experiments.

5.2.5. Assignment of the Lys154Asn α_1 -antitrypsin using NMR spectroscopy

The ^1H - ^{15}N TROSY-HSQC backbone resonance spectrum of Lys154Asn α_1 -antitrypsin at 25 °C was of high quality and generally corresponded very well with that recently assigned for the wild-type protein¹⁴⁰ (Fig.5.15.).

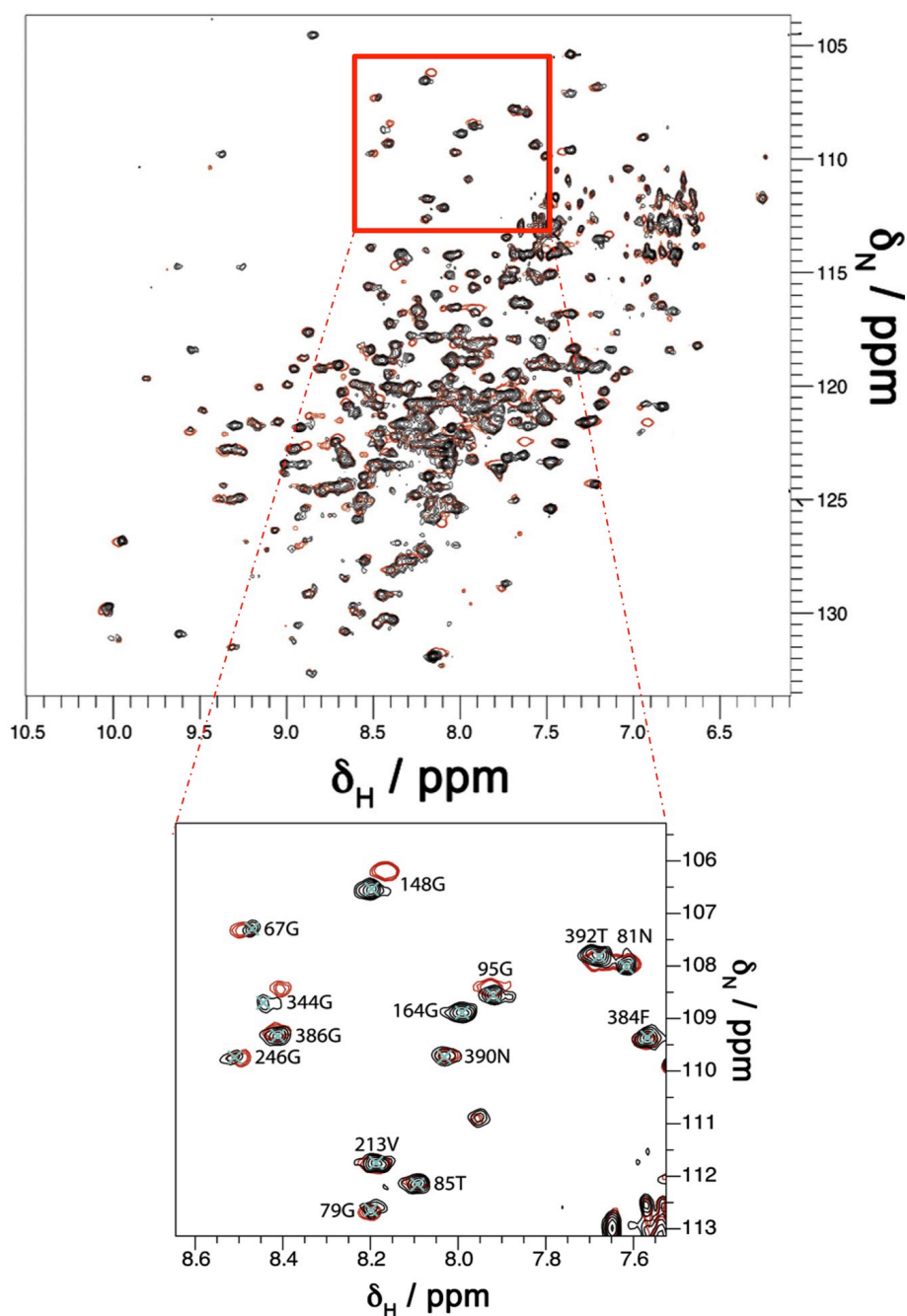


Figure 5.15. ^1H - ^{15}N TROSY-HSQC Spectrum of Wild-Type and Lys154Asn α_1 -antitrypsin. Zoom (right) illustrates examples of reporter residues (wild-type, black with cyan assignment; Lys154Asn, red). The excellent overlay between the wild-type and the Lys154Asn allows the transfer of the assignment for further NMR studies.

The assignments could therefore be unambiguously transferred for 123 resonances that were well-dispersed in the 2D spectra allowing further comparisons to be made at 37 °C as below. The total number of cross-peaks in the dispersed regions were similar in the wildtype and mutant proteins at 25 °C (152 and 148 respectively), as was the combined intensity of crosspeaks in the

central region. The inability to transfer 29 assignments of well-dispersed crosspeaks from the ^1H - ^{15}N TROSY HSQC backbone spectrum of wild-type α_1 -antitrypsin to that of Lys154Asn α_1 -antitrypsin was likely related to the magnitude of the change in chemical shift for these cross-peaks in the Lys154Asn mutant rather than a major change in local dynamics. The 123 comparable cross-peaks were analysed in terms of signal intensity and chemical shift similarities between wild-type and Lys154Asn α_1 -antitrypsin. The degree of variation reported by the relative intensity and chemical shift changes is represented in figure 5.16. The magnitude of changes is shown by heatmap colouring of spheres mapped to the 1.8 Å crystal structure of native α_1 -antitrypsin. These data indicate that, in the native state ensemble, the mutation most affects residues around the F-helix, strands 1-3 of β -sheet A and the reactive loop.

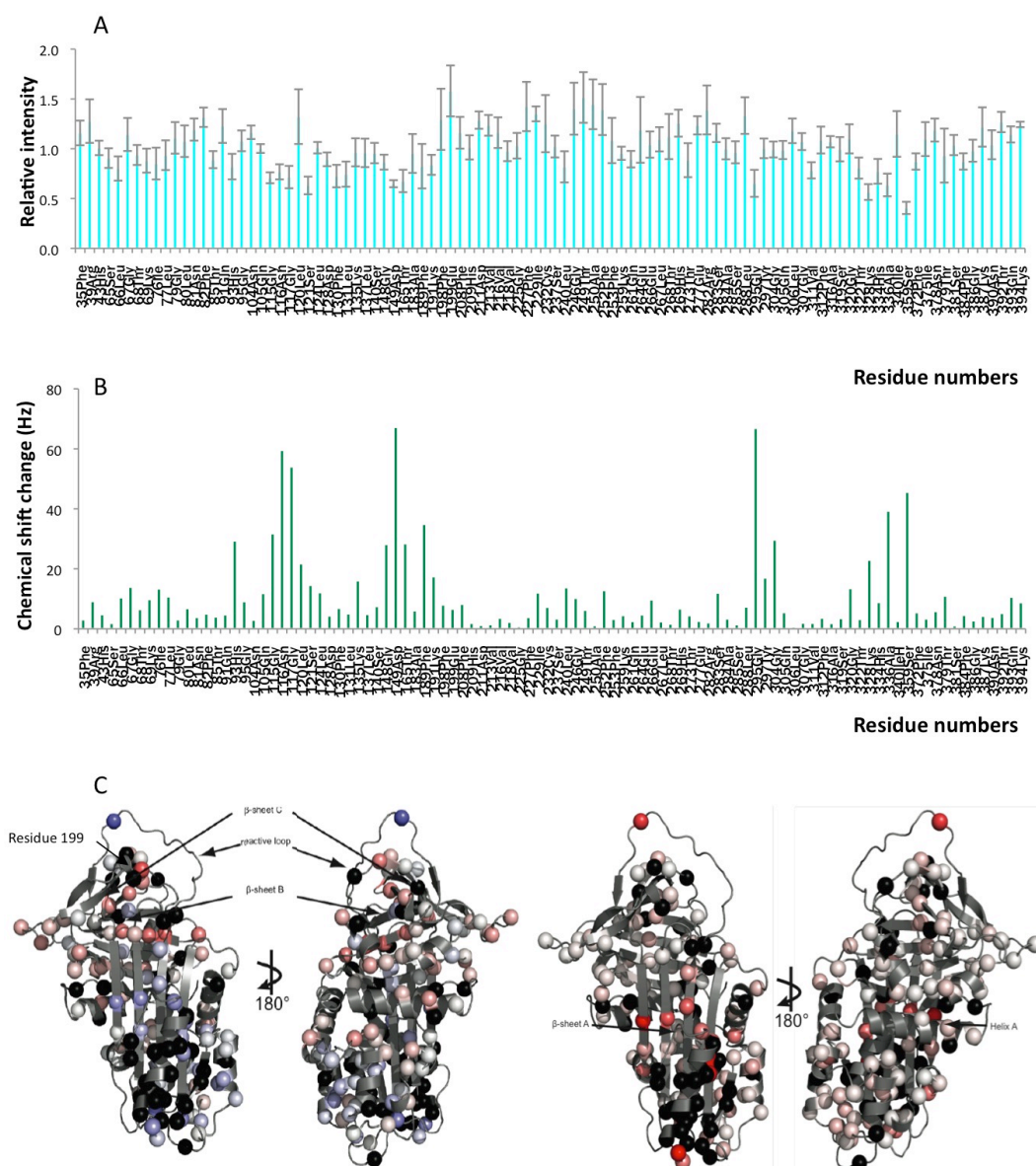


Figure 5.16. Solution behaviour and residue-specific changes induced by the Lys154Asn mutation in α_1 -antitrypsin. A- Changes in cross-peak intensity and B- magnitude of cross-peak chemical shift ($\Delta\delta$, below) for Lys154Asn relative to wild-type α_1 -antitrypsin at 25 °C for the subset of unambiguous resonances in 2D spectra. Error bars are defined according to the standard deviation observed for peak intensities across the wild-type (WT) or Lys154Asn (K154N) datasets at 25 °C using the formula: $[(I_{K154N}/I_{WT})] \times \sqrt{[(d_{K154N}/I_{K154N})^2 + (d_{WT}/I_{WT})^2]}$ where I is cross-peak intensity and d is the standard deviation. C- Residues corresponding to cross-peaks that could not be followed are shown in black. Other residues are coloured grey. Relative intensity (left) data shown plotted on a Red-white-blue heatmap colouring (increases - red, no change - white, decreases - blue). The greatest change was an increase in relative intensity to 1.58 (residue 199): represented by maximal redness on an RGB colour scale. Chemical shift (right) data all changes were calculated as magnitudes and were therefore positive for chemical shift data. The maximum change that could be assigned to a specific residue (0.169 ppm), occurred at residue 149 and so this is represented by maximal RGB scale redness.

5.2.6. NMR of polymerogenic intermediate ensemble at 37 °C

^1H - ^{15}N TROSY-HSQC spectra reported on the formation of the pathological intermediate at 37 °C by non-uniform broadening of previously assigned cross-peaks in Lys154Asn α_1 -antitrypsin relative to wild-type α_1 -antitrypsin. This is shown in Fig. 5.17. by residue-specific intensity change across the primary sequence. This was associated with the appearance of a further 12 discrete cross-peaks in the well-dispersed region of the spectrum in Lys154Asn α_1 -antitrypsin, as illustrated by some examples in figure 5.17.

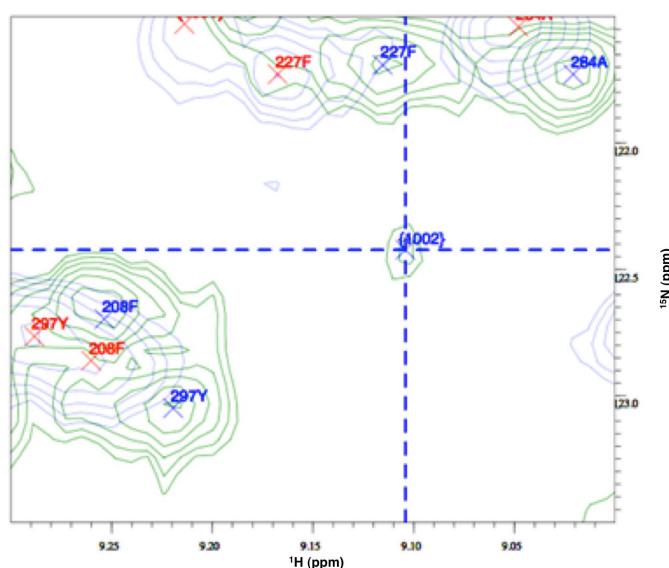


Figure 5.17. ^1H - ^{15}N TROSY-HSQC Spectrum of Wild-Type and Lys154Asn α_1 -antitrypsin. ^1H - ^{15}N TROSY-HSQC of Lys154Asn α_1 -antitrypsin at 25 °C (grey contours, blue labels) and 37 °C (green contours, red labels). Example of 'extra' cross-peak appearing at 37 °C in Lys154Asn α_1 -antitrypsin but not seen at 25 °C or in wild-type α_1 -antitrypsin at either temperature is highlighted (dashed intersection).

There was no loss in the summed cross-peak intensity observed in the central region of the spectrum relative to the wildtype protein at 37 °C. These data are consistent with a process of native \rightarrow intermediate conformational exchange that is slow (milliseconds to seconds) relative to NMR timescales. The mean intensity of the extra peaks was roughly half (0.47) that of the mean intensity of all other discrete cross-peaks in the spectrum, suggesting an approximate 2:1 ratio of native:intermediate conformers of Lys154Asn α_1 -antitrypsin at 37 °C (Fig.5.18.A).

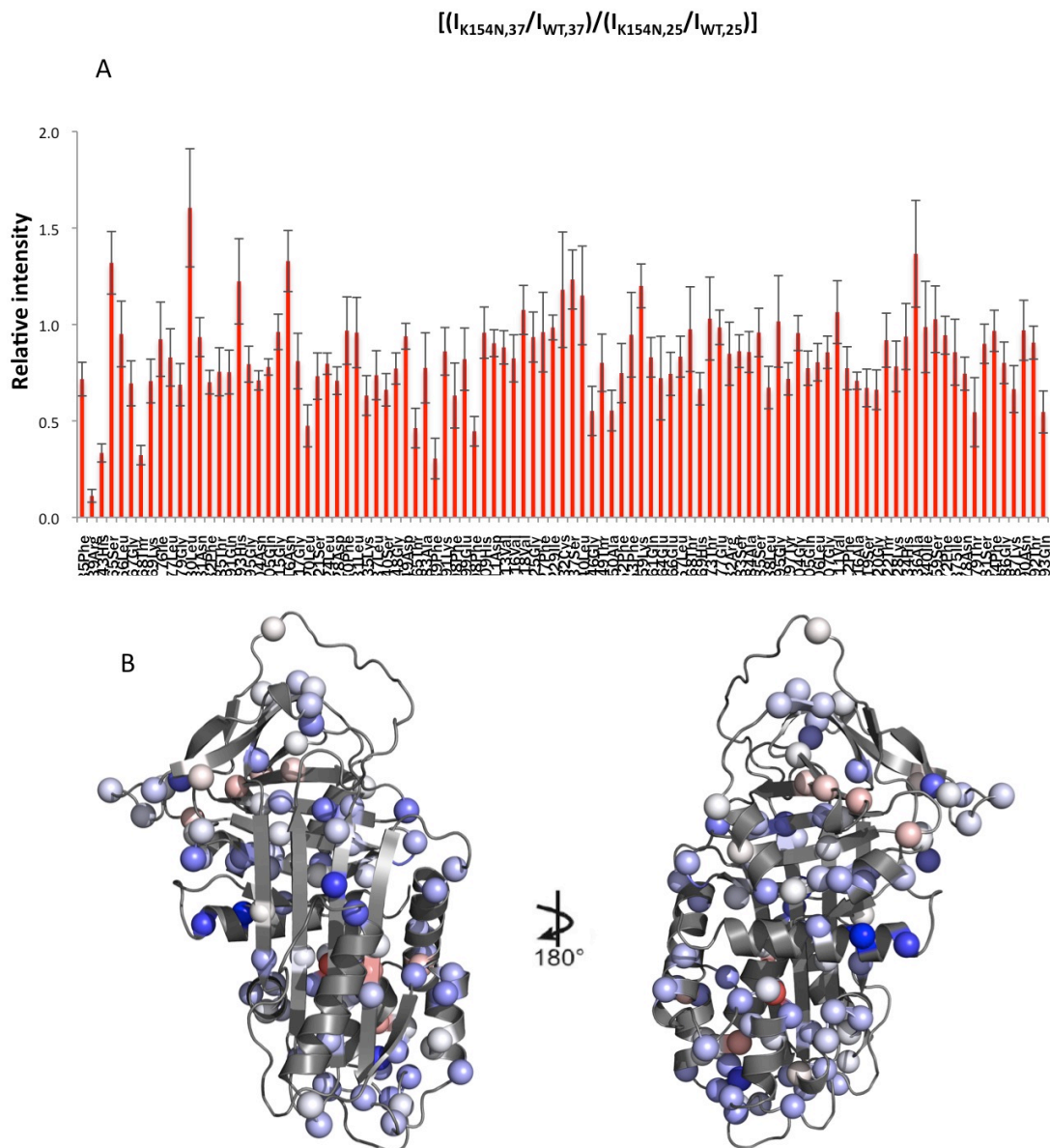


Figure 5.18. Population of the Intermediate by Lys154Asn α_1 -antitrypsin under Physiological Conditions. A- Differential change in intensity for reporter cross-peaks in the ^1H - ^{15}N TROSY-HSQC of Lys154Asn α_1 -antitrypsin at 37 °C. Intensities are scaled relative to the wildtype protein at the same temperature and the intensities observed for Lys154Asn α_1 -antitrypsin relative to the wild-type at 25 °C $\{[I_{K154N,37}/I_{WT,37}]/[I_{K154N,25}/I_{WT,25}]\}$. B- Relative intensities quantified mapped onto the subset of reporter residues (spheres) in red-white-blue heatmap colouring (increases, red; unchanged, white; reductions, blue). Greatest change (maximally blue by RGB, residue 39) corresponds to a relative intensity of 0.11. Increasing redness indicates increasing intensity on the same scale. Ellipse: hA region reporting major change from native-like intensity.

Values are scaled for monomer concentration; they take into account changes observed between wild-type and mutant protein at 25 °C in the absence of polymerisation and changes observed in wild-type α_1 -antitrypsin between 25 and 37 °C. They therefore report differential changes in Lys154Asn α_1 -antitrypsin associated with population of the polymerogenic state. Most cross-peaks that remained unambiguously assigned in Lys154Asn α_1 -antitrypsin reported minimal change at 37 °C relative to wild-type protein and compared with observations at 25 °C (Fig.5.18.B). Most reporter cross-peaks signalled minimal change in Lys154Asn α_1 -antitrypsin at 37°C. NMR intensity data are highly sensitive reporters of changes in structure and dynamics, so relatively large changes may be induced by quite small changes in solution behaviour. A value of 100% relative intensity would therefore indicate entirely native-like structural and dynamical behaviour in a particular residue. Nevertheless residues reporting the most dramatic intensity change in the polymerogenic ensemble notably occur in regions previously linked with conformational change (upper s3A, upper hF, underlying the upper s4A site, hA). However Phe189 in the upper part of strand 3 and the nearby Thr165 at the top of F-helix showed a clear decrease in intensity. Changes of similar magnitude were also seen in β -sheet B where it underlies the breach (upper s4A) region and, interestingly, at longer range around the site of another polymerogenic mutation (I39C) in the A-helix. Mutations in contiguous residues in α_1 -antitrypsin and antithrombin are also associated with deficiency and disease. Our data suggest that structural changes here are integral to formation of the pathological intermediate. These effects may be transmitted most directly to the A-helix from s3A via the intervening shutter and s6B motif.

Residues in strand 5 in β -sheet A, helix I and the intervening linker, which were described to expand during intermediate formation in the β -hairpin model, showed strongly preserved intensity, therefore the data presented here using Lys154Asn α_1 -antitrypsin mild-disease mutant do not support this model of polymerisation. Similarly residues in strands 4 and 5 of β -sheet B, that must unfold in the triple-strand polymerisation model, are also among the more stable residues assessed in this way.

5.3 Discussion

Studying key native to intermediate transitions can elucidate mechanisms of the conformational diseases and aid targeting of novel therapeutic strategies. However the intrinsic tendency of intermediates to adopt more stable, polydisperse, multimeric conformations renders this inherently challenging. Moreover, structurally distinct serpin polymers are induced by chemical denaturants compared to pathological mutations or heating⁹³.

The Lys154Asn mutant of α_1 -antitrypsin (α_1 -antitrypsin Queen's) was identified in an individual with deficiency of the plasma protein. The mutation results in the formation of polymers *in vitro* and when expressed in a eukaryotic cell model of disease. Lys154Asn α_1 -antitrypsin polymers induced by heating or expression within the cell model are recognised by the 2C1 monoclonal antibody (mAb). 2C1 recognizes disease polymers of Z α_1 -antitrypsin that form *in vivo* and when induced *in vitro* by heating. It is highly conformer-specific and can even discriminate these polymers from those formed in the presence of denaturant^{93; 189}. Accordingly, data obtained for Lys154Asn α_1 -antitrypsin are likely to be applicable to other pathological, polymerogenic mutants of α_1 -antitrypsin.

Furthermore the Queen's variant adopts the native conformation at 25 °C but readily populates a polymerogenic intermediate ensemble in physiological conditions. However it forms polymers far more slowly than Z α_1 -antitrypsin. This combination explains the low level of polymerisation within the cell model and the relatively mild deficiency of circulating protein and lack of clinical disease in the index case (Queen's/Z compound heterozygote). Critically the same combination of characteristics rendered the polymerogenic intermediate ensemble unusually amenable to biophysical and biochemical characterization, including the use of NMR spectroscopy. These studies report at the level of multiple specific residues located all around the protein, representing > 30 % of the entire sequence. Thus they allow the first detailed characterization of the structural and/or dynamic solution behaviour of α_1 -antitrypsin during formation of a disease-relevant polymerogenic intermediate ensemble.

Wild-type and Lys154Asn α_1 -antitrypsin demonstrated similar peak counts in ^1H - ^{15}N TROSY-HSQC NMR spectra at 25 °C, consistent with IM-MS data suggesting a similarly high population of the native state in the two proteins in these conditions. However at 37 °C the data were consistent with increased population of the intermediate in Lys154Asn α_1 - antitrypsin compared with the wildtype protein. The findings were in good agreement with the data from biochemical, biophysical and cellular studies across physiological temperatures. Intermediate formation was therefore assessed by comparing spectra of wildtype and Lys154Asn α_1 -antitrypsin at both 25 °C and 37 °C. Residue-specific changes in intensity were quantified by ratios that accounted for changes in native signal reporting due to the mutation and the increase in temperature, and a low constant loss of global signal due to background polymerisation. The changes that are observed by this approach therefore solely report structural and/or dynamic changes affecting individual residues due to the population of the polymerogenic intermediate ensemble.

The average relative intensity across the entire subset of residues that could be analysed in this way was 82%. It is therefore unlikely that the intermediate state is substantially unfolded as required for extensively proposed domain-swapped models of polymerisation. Moreover, analysis of cross-peaks indicates widespread preservation of highly native-like solution behaviour over most of the molecule. In particular the data support high stability in the hI-s5A and s4B, s5B regions proposed to unfold in the β -hairpin and triple-strand models of polymerisation respectively ^{97; 99}. Instead intermediate ensemble formation was characterized by highly-localized changes in structural and/or dynamic behaviour. These changes affected residues in regions associated with remodeling during intermediate ensemble formation in the single strand model of polymerisation (Thr165 at the top of F-helix, relative intensity 0.46, Phe189, upper s3a, 0.31). Moreover, a major loss of relative intensities was noted at the C-terminal end of the A-helix (relative intensity: Arg39, 0.11 and His43, 0.33). The F-helix and linker region may therefore constitute a “clasp” motif, sealing a network of interactions that regulate conformational change. Similar clasp

unlocking likely underlies the polymerogenic effects of antithrombin Rouen-VI (Asn158Asp by α_1 -antitrypsin template numbering) mutation in the related serpin antithrombin that is also associated with deficiency²⁰⁰. On the other hand, a linker mutation that promotes polar interactions within the clasp (Asn171His, template numbering) stabilises the native serpin plasminogen activator inhibitor-1 (PAI-1) against loop insertion and conformational change²⁰³. These data from this particular α_1 -antitrypsin variant are most readily reconciled to the classical single-strand polymerisation model⁷⁷. However, further NMR studies are required for definitive characterisation of which polymerisation model(s) is/are most relevant to disease *in vivo*.

To conclude, this study illustrates arguably the best strategy to study formation of the pathological intermediate. It shows how powerful, but non-perturbing methods can characterise the solution behaviour of *forme fruste* variants in physiological conditions providing insights that can be related to more severe mutants.

Chapter **6** - DISCUSSION

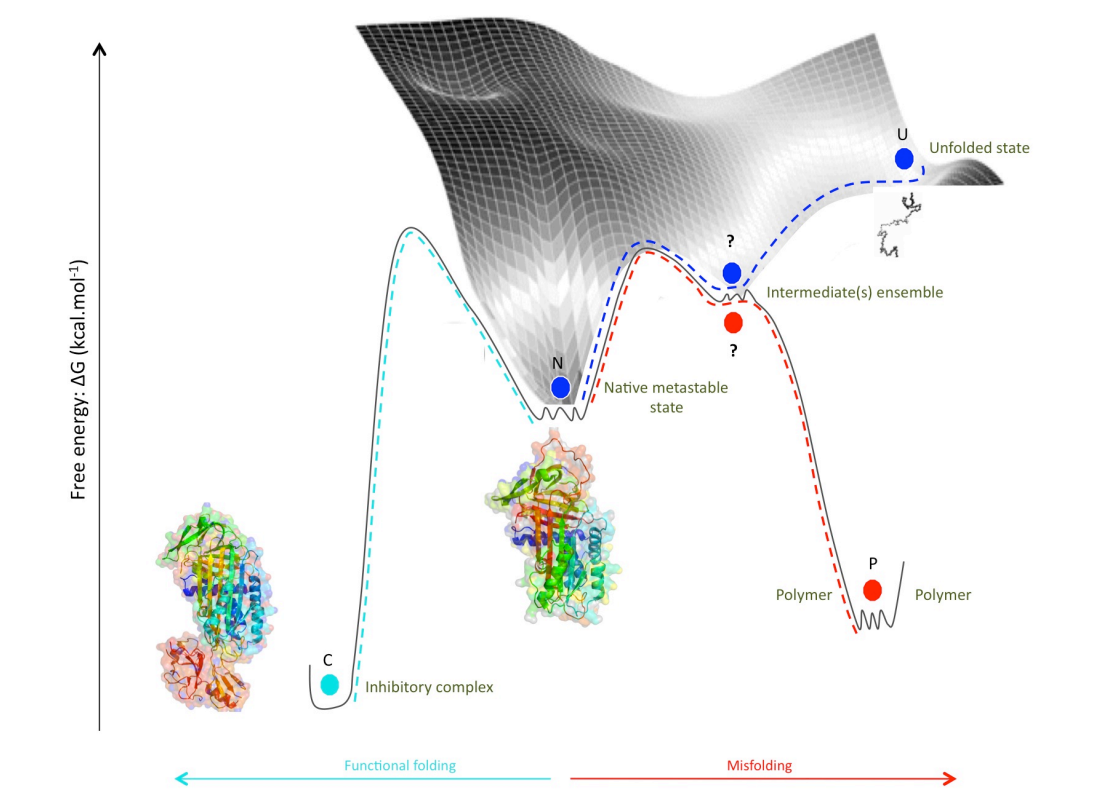


Figure 6.1. Energy free-energy landscape α_1 -antitrypsin. Protein energy landscapes are typically many-dimensional, defining more or less favoured pathways between distinct conformational states. This figure schematically illustrates such a landscape for α_1 -antitrypsin. The putative unfolded to folded transition is shown in blue, with the metastable native state shown as a local energy minimum. Binding to its target proteinase results in a highly stable inhibitory complex (cyan). Furthermore, genetic mutations observed in disease lead to misfolding and aggregation to highly stable polymers (red) (also shown in section 1.2.2, Chapter I).

This thesis puts forward a strategy for the use of Nuclear Magnetic Resonance spectroscopy together with complementary biophysical methods, to unveil processes of folding, misfolding and polymerisation of α_1 -antitrypsin in unprecedented detail. These processes underlie function and dysfunction in the serpin superfamily of proteins.

Human diseases characterized by protein conformational modifications have been recognized for almost two centuries^{47; 53; 204; 205; 206}. It has now become clear that many different proteins can misfold and form extracellular, or intracellular aggregates that initiate profound cellular dysfunction²⁰⁷. Particularly challenging examples of such disorders include Alzheimer's and Parkinson's diseases. Another predominant pathology characterized by the ordered aggregation and tissue deposition of aberrant protein conformation, described in this work, are the serpinopathies²⁰⁸.

As well as for most conformational diseases, the structure and the formation of pathogenic polymers remains subject of debate. For the serpinopathies this is of particular importance. They differ from other conformational diseases where fibrillar deposited states may be less toxic than smaller soluble species in that deposition of polymeric species appears to tightly correlate with disease severity²⁰⁹. This implies that the pathogenic state(s) are likely to occur on the pathway to polymerisation and that the polymers themselves are toxic. The presence of a polymerogenic intermediate ensemble in the pathway, often denoted as M^* , is supported from the kinetics of polymerisation of serpins^{81; 156; 210}. Structural data and studies with the α_1 -antitrypsin polymer specific monoclonal antibody 2C1 support the population of multiple intermediate states in a condition-dependent manner^{77 93;208 97; 99 115 189}. *In vitro* polymerisation therefore appears not to be a single phenomenon, but to proceed over a complex energy landscape incorporating a range of ensemble conformations and multiple possible end products. This landscape also relates to a physiological folding pathway that avoids populating the hyperstable polymeric and monomeric conformations in favour of the metastable and hence functional native state (outlined in simplified form in Fig.6.1.).

NMR spectroscopy can report upon structural and dynamic solution behaviour over time scales covering 6-9 orders of magnitude^{128; 130}. On the other hand, applying these methods to a protein of the size of the prototypic serpin α_1 -antitrypsin is challenging even after the 'first step' of assigning ~90% of its polypeptide chain residues in the backbone NH spectrum¹⁴⁰. Nevertheless this breakthrough allows us to ask more sophisticated questions by a range of NMR techniques, to interrogate how the motions and dynamic properties of individual residues within α_1 -antitrypsin may facilitate conformational change. The next steps that I have undertaken in this thesis, aided by complementary biophysical techniques, have primarily been to characterise:

- i) transitions observed in urea denaturing equilibrium studies to inform on folding and misfolding processes;
- ii) the dynamics of the natively folded metastable state in solution using HDXNMR (ms-s timescale) in order to understand how potential lability is regulated;
- iii) Polymerogenic intermediate ensemble formation in physiological conditions in a novel variant associated with deficiency.

Various biochemical and biophysical techniques have been used to separate and characterize proteins with different oligomerization states and/or conformations formed by perturbation of the native state^{211; 212; 213}. Examples of such perturbations are those induced by changes in solution conditions, such as temperature or pH. Alternatively, denaturant studies are widely used to probe protein folding and unfolding pathways¹⁶¹. They have been considered particularly relevant in studying the relationship of these processes to disease-associated polymerisation in systems, where low concentrations of denaturant, can induce the formation of polymeric states^{99; 161}.

Over 20 years ago, the first published study of α_1 -antitrypsin unfolding described equilibrium studies across a guanidinium hydrochloride titration¹⁶⁵. It has since led to numbers of biochemical and biophysical unfolding studies using both urea and guanidine^{83;166;81; 100; 167}. These have suggested that the folding/unfolding pathways of α_1 -antitrypsin may involve three states: native

(N), intermediate (I), unfolded (U) ^{83;166;81; 100; 167} with intermediate states maximally populated around 1.5 M guanidine⁸³ or 4.0 M urea⁸⁸. In order to obtain residue-specific information on conformational change during such processes, I undertook a urea titration from 0 M to 8.0 M urea (100 μ M α_1 -antitrypsin), and followed this by 2D (¹H-¹⁵N TROSY-HSQC) NMR spectroscopy experiments (Chapter III, Fig.6.1. blue pathway). This strategy was combined with 1D (¹H) NMR and CD spectroscopy, in order to track off-pathway polymerisation at the high protein concentration and global unfolding when such polymerisation was minimized (0.1 mg/ml α_1 -antitrypsin), respectively. Spectra at 0 M and 8.0 M urea demonstrated the characteristic folded and unfolded fingerprints of α_1 -antitrypsin, respectively. Surprisingly, at low-urea concentration, 1.0 M urea, where the native conformation appears stable by CD spectroscopy conformational changes were already observed in the solution behaviour of α_1 -antitrypsin. These may therefore predominantly reflect changes in dynamic behaviour rather than structural remodelling. Intriguingly, a few residues in β -sheet C reported enhanced 'native-like' signal in these conditions. A possible explanation is that these α_1 -antitrypsin residues are less conformationally labile at 1.0 M than at 0 M urea and could act as a more rigid scaffold during late folding/early unfolding.

All dispersed resonance signals were lost during equilibration of 100 μ M α_1 -antitrypsin at 2.0 M urea. Whilst the corresponding CD equilibrium studies at \sim 50-fold lower α_1 -antitrypsin concentration did not support extensive unfolding. Since unfolding is not likely to be concentration-dependent this result indicates that the combination of denaturant and high α_1 -antitrypsin concentration is causing urea-induced polymerisation of most of the sample in the conditions used for the NMR studies. Urea-induced polymerisation was dominant in the 2.0 to 6.0 (maximal 3.0-4.0) M urea range. However, the data do not demonstrate clear evidence of a distinct urea-induced intermediate. Unfolded resonances started appearing between 3.0 to 7.6 M urea.

Unexpectedly, CD spectral studies strongly supported a two-states equilibrium unfolding behaviour. The absence of an apparent intermediate state definable by a distinct secondary structure signature in urea is surprising.

Previous studies of α_1 -antitrypsin by single wavelength CD spectroscopy (ellipticity reported at 222 nm) supported the population of an extra state at 2.5 - 4.0 M urea^{88;83}. My studies suggest this 'state' may have been an artefact of polymerisation induced by urea when α_1 -antitrypsin was > 0.25 mg/ml. My CD studies were undertaken at 0.1 mg/ml α_1 -antitrypsin, where polymerisation could not be detected by other methods. This low protein concentration was inevitably associated with noisier spectra than would be the case with higher α_1 -antitrypsin concentrations. However, it effectively minimizes the effects of polymerisation that is otherwise favoured for this range of urea conditions. The findings presented in this thesis do not exclude the possibility that in other environmental conditions (e.g. low-pH, guanidine, heat) unfolding pathways follow a three-state process. They do however, constitute a useful comparator for further unfolding exploration. With the exception of guanidine, these could be investigated further at residue-specific level and in solution, using NMR spectroscopy. They also indicate the importance of experimentally validating 'intermediate' serpin states observed in denaturing conditions by assessing their apparent signal at a range of protein concentrations.

Furthermore, the severe Z disease mutation (Glu342Lys) results in only mild reductions in thermodynamic stability relative to the effects of stabilizing mutations (Thr114Phe and Gly117Phe). This is important since, despite the larger thermodynamic effects, these phenylalanine mutations themselves cause no or minimal rescue of the disease phenotype when introduced on a Z α_1 -antitrypsin background.

The study of the dynamics of the metastable state presented in this work, started with exploration of the motions in the range of the milliseconds to seconds timescale using HDXNMR experiments (chapter IV). Indeed, HDXNMR was previously described to constitute a powerful tool providing insight into equilibrium populated intermediate²¹⁴. The true challenge of this strategy lies on both the size of the protein, and the rapidity of exchange for many labile residues and the tendency to undergo acid-induced polymerisation below pH 6.0. These were addressed by optimizing the system (protein concentration, temperature, type of NMR experiment, method of exchange, length of

experiment), working at high concentrations of α_1 -antitrypsin (250 μ M), using SOFAST Heteronuclear Multiple Quantum Correlation¹³⁸ (HMQC) NMR experiments that allow rapid monitoring of fast exchange²¹⁵, and by assessing the sample for polymerisation during and after the experiments. When assessed by HDX-NMR, the data demonstrated groupings of residues with high rigidity or high lability. Relatively few residues reporting intermediate levels of protection on this scale. Globally, the lability ranged between $0.4 \text{ kcal.mol}^{-1} \geq \Delta G_{\text{app}} \leq 9 \text{ kcal.mol}^{-1}$. This range and the distribution of values is consistent with previous data on global ΔG values⁸¹ and earlier lower resolution, higher coverage HDX-MS data⁹² to which it is complementary. More specifically, the data define a region of high protection extending from β -sheet C through the core (β -sheet B and B-helix) through to the middle of β -sheet A that can be divided into 2 hydrophobic clusters. Cluster 1 locks closed the centre of β -sheet A (strands 3 and 5) by interaction with the B-helix and the lower, buried strand residues in β -sheet B. It corresponds to the shutter region, and boundaries of the s4A/reactive loop insertion site^{77; 140}. The second cluster affects hydrophobic residues interacting in a β -barrel region between β -sheets B and C. Destabilisation of this region by 'latch mutations' was recently described as a pathogenic mechanism in mild deficiency variants of α_1 -antitrypsin¹⁹¹ and this region is also proposed as a nucleation site for protein folding⁹⁰. The minimally-protected, flexible regions include the RCL and an area around the interface between the F-helix and β -sheet A. This suggests that the dominant regulation of opening is by cluster 1 interactions rather than conformational stability of the RCL or F-helix. The effects of disease mutations support this. Whilst mutations shown or predicted to destabilize these regions do cause polymerisation and circulating deficiency^{192; 193; 194} the effects of destabilizing s5A and shutter region mutations are much greater^{95; 173; 195}. Moreover, I have conducted spin relaxation experiments^{216; 217} whose data are currently being analysed to inform upon fast (ps-ns timescale) dynamics in α_1 -antitrypsin. Building on these methods, the use of relaxation-dispersion NMR^{218; 219} may allow analysis of low-populated intermediate species in solution, e.g. during folding or pathophysiological conditions in the context of disease mutations.

Studying key native to intermediate transitions can elucidate mechanisms of the conformational diseases and aid targeting of novel therapeutic strategies. However the intrinsic tendency of intermediates to adopt more stable, polydisperse, multimeric conformations renders this inherently challenging. Exploring the native to polymer transition (Chapter V, Fig.6.1., red pathway), this thesis presents a newly characterised *forme fruste* deficiency variant of α_1 -antitrypsin (Lys154Asn) that forms polymer recapitulating the conformer-specific neoepitope observed in polymers that form *in vivo* (2C1). In this work, NMR spectroscopy was used to provide new insights on the structural and dynamic changes, and potential formation of M^* , associated with polymer formation. The Lys154Asn mutant of α_1 -antitrypsin (so-called α_1 -antitrypsin Queen's) was identified in an individual with deficiency of the plasma protein. It adopted the native conformation at 25 °C but this thesis shows that it readily populated a polymerogenic intermediate ensemble at pathophysiological temperatures (characterised by Native-PAGE, CD spectroscopy and IM-MS studies). It was shown however to form polymers far more slowly than Z α_1 -antitrypsin. The combination explains the low level of polymerisation within the cell model and the relatively mild deficiency of circulating protein and lack of clinical disease in the index case (Queen's/Z compound heterozygote).

The 'molten globule' state as a general term was introduced in section 1.1.1.3, describing for the first time the existence of a third (intermediate) equilibrium. This was defined as a thermodynamic state that conserves a native-like secondary structure content without the tightly packed protein interior (lacking a specific tertiary structure). The results presented in this thesis provide evidence for different conformational behaviour along pathways of polymerisation induced by denaturants and by disease mutations. This indicates they cause the polypeptide chain to explore different paths through the folding landscape (Fig. 6.2) leading to different stabilized polymer states.

The same combination of characteristics rendered the polymerogenic intermediate ensemble unusually amenable to biophysical and biochemical characterization, including the use of NMR spectroscopy. These studies report at the level of multiple specific residues located all around the protein, representing

> 30 % of the entire sequence. Thus they allow the first detailed characterization of the structural and/or dynamic solution behaviour of α_1 -antitrypsin during formation of a disease-relevant polymerogenic intermediate ensemble. Intermediate formation was therefore assessed by comparing spectra of wild-type and Lys154Asn α_1 -antitrypsin at both 25 °C and 37 °C. The residue-specific changes in intensity that are observed by this approach therefore solely report structural and/or dynamic changes affecting individual residues due to the population of the polymerogenic intermediate ensemble. The data support high stability in the hI-s5A region proposed to unfold in the β -hairpin model of polymerisation⁹⁷. Instead intermediate ensemble formation was characterized by highly localized changes in structural and/or dynamic behaviour. These changes affected residues in regions associated with remodelling during intermediate ensemble formation in the single strand models of polymerisation (top of F-helix, upper s3A). Moreover, a major loss of relative intensities was noted at the C-terminal end of the A-helix. The F-helix and linker region may therefore constitute a 'clasp' motif, sealing a network of interactions that regulate conformational change. Similar clasp unlocking likely underlies the polymerogenic effects of antithrombin Rouen-VI (Asn158Asp by α_1 -antitrypsin template numbering) mutation in the related serpin antithrombin that is also associated with deficiency²⁰⁰. On the other hand, a linker mutation that promotes polar interactions within the clasp (Asn171His, template numbering) stabilises the native serpin plasminogen activator inhibitor-1 (PAI-1) against loop insertion and conformational change²⁰³. Based upon recent experience with the serpinopathies, it seems reasonable to propose that the use of powerful techniques that are non-perturbing to solution equilibria of *forme fruste* disease variants is an optimal strategy for the study of conformational disease processes. Data from this particular α_1 -antitrypsin variant are most readily reconciled to the classical single-strand polymerisation model⁷⁷. However, further NMR studies are required for definitive characterisation of which polymerisation model(s) is/are most relevant to disease *in vivo*.

Overall, my studies support the use of NMR spectroscopy to obtain unique residue-specific structural and dynamical information of the different folding and misfolding/polymerisation pathways available to α_1 -antitrypsin. They also illustrate the importance of corroborating observations from this technique with those from complementary methods to avoid misinterpretation of potentially complex data. This opens the way to future studies including those that can be applied to aid the development of new therapeutic strategies to treat α_1 -antitrypsin deficiency by affecting structural and dynamic behaviour of the protein in solution. The dynamics data obtained on the native state of wild-type α_1 -antitrypsin using HDX-NMR and relaxation experiments allow future comparison with the behaviour of disease mutants such as Lys154Asn over slow and fast timescales. Furthermore, the NMR spectroscopic characteristics of the native protein described in this thesis provide the basis for assessing its interactions with ligands in drug discovery. To these ends I undertook preliminary drug-binding experiments; findings are presented in the supplementary material B (Appendix B Fig. 1, Fig.2). In this study, the binding of the tetrapeptide lead compound Ac-TTAI-NH₂ to α_1 -antitrypsin was assessed by NMR spectroscopy and IM-MS. The data confirm widespread conformational changes upon binding and a 2:1 stoichiometry, respectively. This method can be used to screen compounds identified as potential ligands, to shed light on the binding interactions with the ligands at atomic resolution.

Furthermore, the work presented in this thesis opens ways to explore whether misfolding processes occur co-translationally on the ribosome in disease variants of α_1 -antitrypsin using NMR spectroscopy of the ribosome-nascent chain (RNC) complex. This has the potential to transform our understanding of the formation of intermediate ensembles in health and disease *in vivo* and it has been very exciting being part of the development of this work.

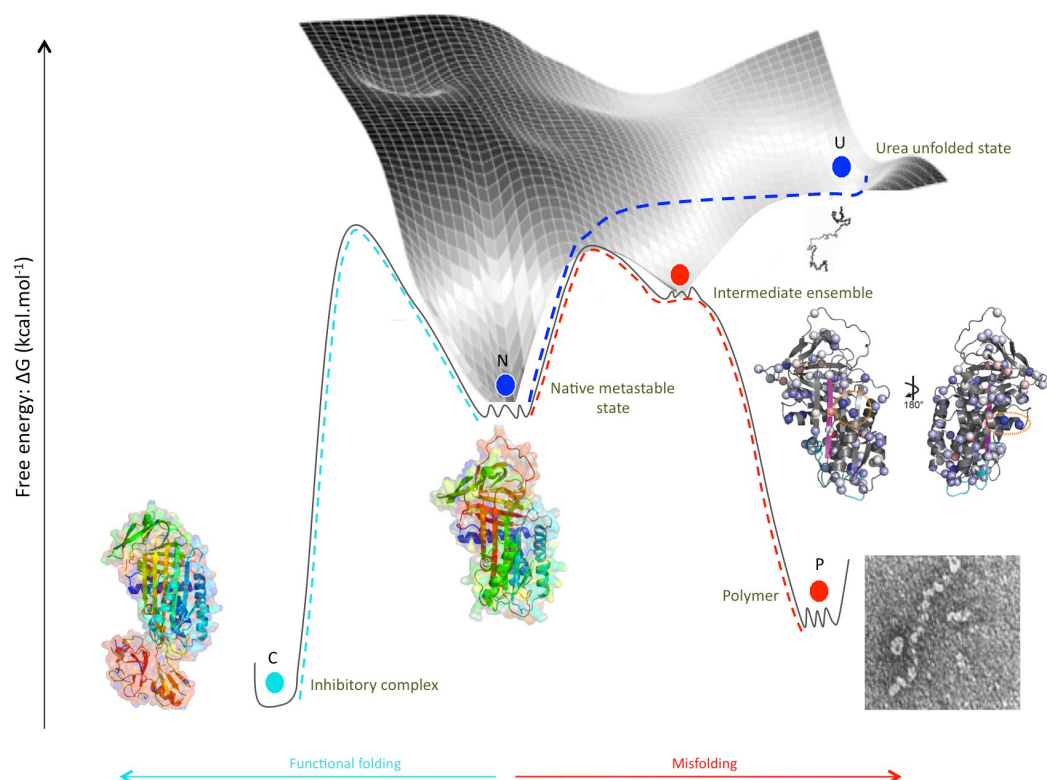


Figure 6.2. Energy free-energy landscape α_1 -antitrypsin revised from the findings of this thesis. Energy landscape shown in Fig.6.1. modified in line with the finding from this thesis, notably the absence of an identifiable, thermodynamically-stabilised intermediate state along the urea-induced folding/unfolding pathway. In contrast the Lys154Asn mutation appears to strongly enrich population of an intermediate ensemble in pathophysiological conditions along a pathway that results in polymers recapitulating the 2C1 neo-epitope.

SUPPLEMENTARY MATERIALS

Appendix A: Fast Timescale Dynamics from NMR relaxation experiments of α_1 -antitrypsin

[work carried out in collaboration with Dr Chris Waudby (Institute of Structural and Molecular Biology, University College of London)]

A.1. Introduction to R_1 , R_2 , and the $^{15}\text{N}\{-^1\text{H}\}$ NOE

The relaxation processes of a perturbed NMR system back to equilibrium occurs through three main mechanisms, longitudinal relaxation (spin-lattice relaxation), transverse relaxation (spin-spin relaxation) and longitudinal cross-relaxation (the Overhauser effect) ²²⁰. These rates are dependent on the fluctuating magnetic environment within the molecule and, in the particular case of ^{15}N nuclear spin relaxation, by the motion of the N-H amide bond vector. Their measurement therefore provides important information on the rotational diffusion of the molecule, and the extent of local mobility and conformational exchange.

Longitudinal relaxation rates may be recorded via inversion recovery experiments (or similar)²²¹. These experiments involves flipping the bulk magnetisation on to the -z axis and allowing the system to evolve for a defined period of time during which the system relaxes towards its equilibrium position; rotating the remaining magnetisation into the x-y plane then permits the detection of the magnetisation as a function of the time allowed for relaxation, allowing the determination of the longitudinal relaxation rate, R_1 , that governs this process. As longitudinal relaxation is driven by spin flips, the observed R_1 rate depends on fluctuations at the Larmor frequencies of ^1H and ^{15}N nuclei and therefore is highly sensitive to the rotational correlation time of the molecule and to molecular motion on ps–ns timescales. ²²¹

Transverse relaxation rates can be recorded using a CPMG experiment. In this type of experiment, the magnetisation is flipped to the -y axis with a 90°_x pulse. The system is then allowed to relax for a period (τ) prior to 180°_y pulse.

The system then relaxes for a time (2τ), followed by another 180°_y pulse, and allowed to relax for a time (τ) prior to acquisition. In practice the CPMG sequence ($\tau - 180^\circ_y - 2\tau - 180^\circ_y - \tau$) is often repeated several times. Varying the time for which the system relaxes allows quantification of the rate of relaxation delay and hence the transverse relaxation rate R_2 . This rate, like R_1 , depends on the overall rotational correlation time of the molecule, but importantly the transverse relaxation rate is also uniquely sensitive to chemical exchange on ms timescales. Alternatively, R_2 relaxation rates may be deduced from rotating frame $R_{1\rho}$ relaxation rates measuring during application of a spin-lock, after correction for longitudinal relaxation in the tilted frame:

$$R_2 = R_{1\rho} / \sin^2 \theta - R_1 / \tan^2 \theta$$

where $\tan \theta = \omega_1 / \Omega$ is the angle of the tilted frame relative to the static field, Ω is the resonance offset and ω_1 is the spin-lock field strength.

The $^{15}\text{N}\{-^1\text{H}\}$ heteronuclear NOE experiment involves the interleaved recording of ^1H - ^{15}N correlation spectra with and without irradiation of ^1H nuclei^{222; 223}. Irradiation of the ^1H spins equalises the ^1H spin population between the higher and lower energy states, and dipolar cross-relaxation (a through-space interaction also known as the nuclear Overhauser effect) between ^1H and its attached ^{15}N then may result in a difference in the intensity of the ^{15}N resonance relative to the experiment conducted without irradiation. The observed intensity difference, termed the steady-state NOE, arises from competition between the cross-relaxation and longitudinal relaxation rates, and is most sensitive to fluctuations at the proton Larmor frequency (of the order of picoseconds). Thus, the steady-state NOE is an excellent probe of local flexibility on the fastest timescales.

A.2. Data acquired and analysis to date:

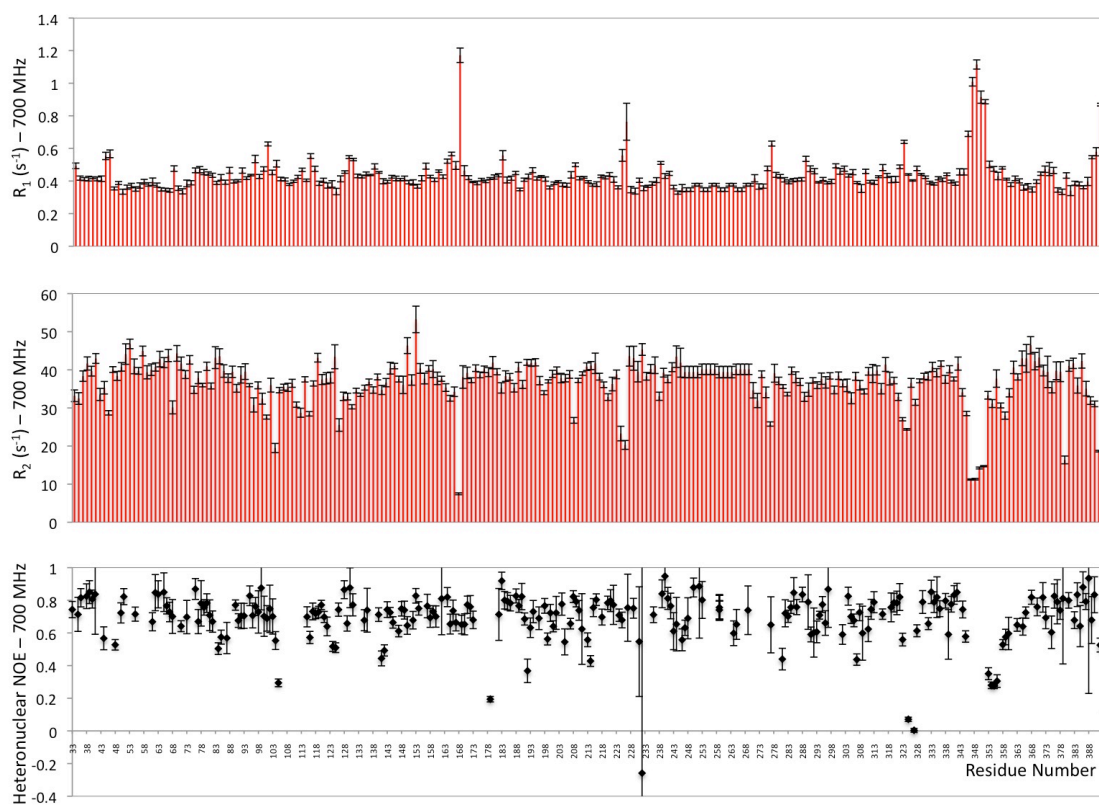
A.2.2. Methods

To investigate the dynamic properties of α_1 -antitrypsin, a series of ^{15}N R_1 , $R_{1\rho}$ and NOE nuclear spin relaxation experiments were recorded by Dr Chris Waudby on a 150 μM sample of perdeuterated, amide-protonated, ^{15}N -labelled α_1 -antitrypsin. TROSY R_1 , $R_{1\rho}$ and NOE experiments were carried out with pulse sequences as described in Lakomek et al²²⁴. Spectra were recorded at 298 K at 700 MHz (see Chapter II of main thesis) Bruker Avance III NMR spectrometer, equipped with a TXI cryogenic probehead and a z-gradient accessory. ^{15}N and ^1H spectral widths were 2128 Hz and 10504 Hz, with 1024 and 160 complex points in each dimension respectively. For each relaxation time measurement, the relaxation decay was sampled for 10 different delay durations, interleaved at the level of single scans. Longitudinal (R_1) relaxation rates were measured using ten points with delays of 0, 0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2 and 3.6 sec. Rotating frame ($R_{1\rho}$) relaxation rates were measured in a 2.05 kHz spin-lock field with relaxation delays of 1, 5, 10, 15, 20, 25, 30, 40, 60 and 80 ms. The spin-lock field strength was measured directly by a nutation experiment, and the observed B_1 -field inhomogeneity was approximately 50 Hz. For temperature compensation purposes an off-resonance spin-lock pulse was applied such that the total duration of the spin-lock was 80 ms, in each scan, for all relaxation measurements. RF carriers were at 4.79 ppm for ^1H and at 118 ppm for ^{15}N . A recycle delay of 2.5 s was used in both R_1 and $R_{1\rho}$ measurements, chosen relative to the measured average ^1H T_1 value of 1.6 s. Heteronuclear NOE measurements were performed with a saturation period of 5 s following a recycle delay of 1 s (or 6 s in the absence of the saturation pulse chain), with saturation pulses applied to the centre of the amide region at 8.5 ppm. Data were processed with Lorentz-to-Gauss window functions and were fitted using nmrPipe¹⁴⁴.

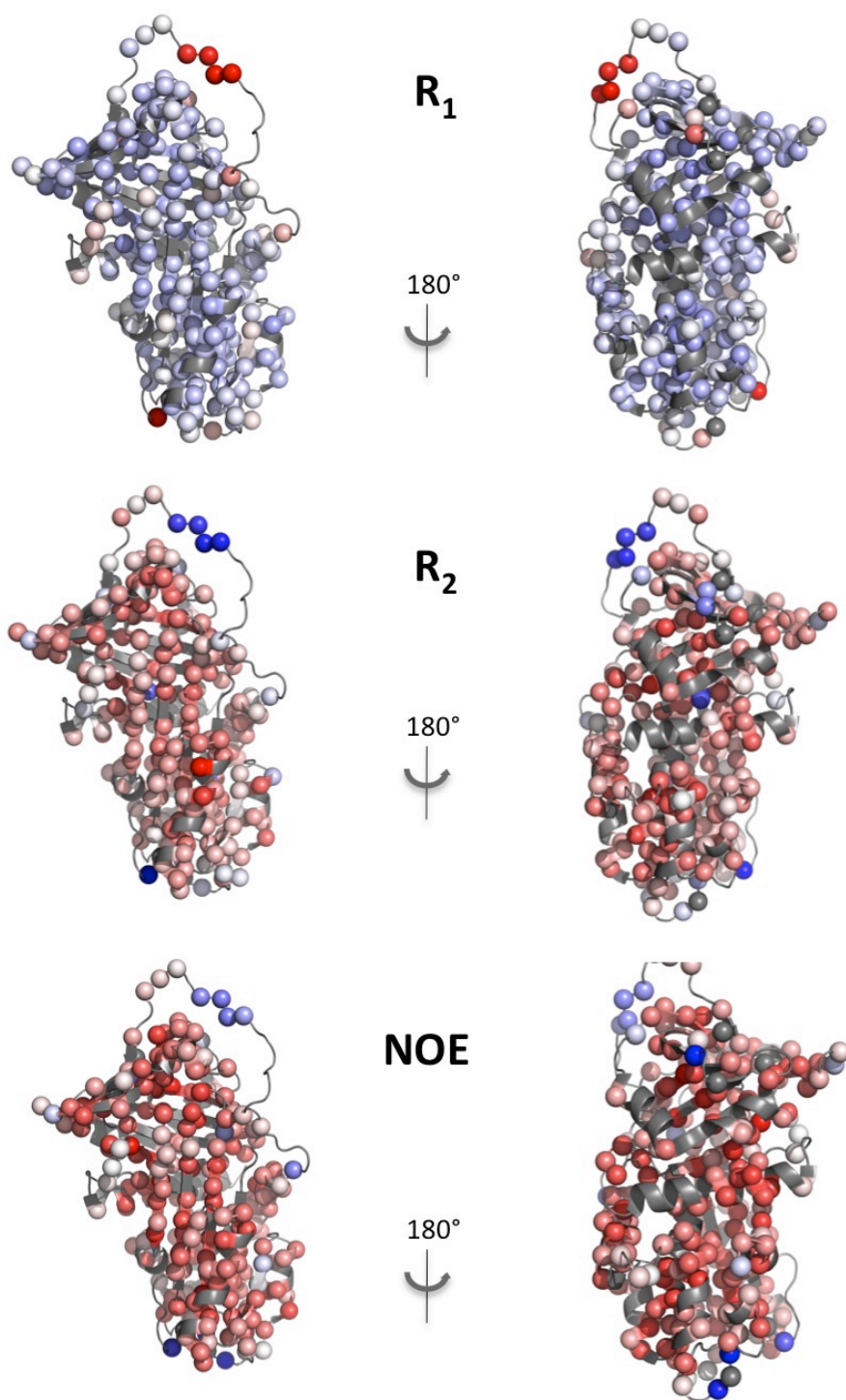
A relaxation analysis (by Dr. Chris Waudby, UCL) was performed for 247 residues out of a possible 394 residues. A preliminary analysis of these data is

presented here and an analysis including additional data recorded by Dr Waudby at 900 MHz is currently underway. This will include a 'Model Free' analysis^{225 226} and a detailed definition of the spectral density function for α_1 -antitrypsin. The parameters determined within the 'model-free' analysis include the diffusion tensor, describing the rotational correlation time(s) (τ_c), generalised order parameters (S^2) and chemical exchange terms (R_{ex}), and in favourable cases order parameters can also be determined for molecular motion on two different time-scales (S_f^2 and S_s^2).²²⁷ Of particular relevance here is the R_{ex} term since this is essentially part of the measured R_2 term, describing the additional contribution to R_2 arising from 'chemical exchange'. The careful analysis of such exchange terms, using so-called 'relaxation dispersion' CPMG or $R_{1\rho}$ measurements, is forming a very important part of the use of NMR in biology as it provides a route to the high-resolution characterisation of intermediate states^{122 228 229} that are weakly populated at equilibrium, but that are in exchange with the ground state on ms timescales. This is of particular interest in the case of α_1 -antitrypsin as sparsely populated intermediate states that are to date very poorly understood may be of relevance to the misfolding processes that occur in disease²³⁰. However, this type of analysis is unfortunately beyond the scope of this thesis as it requires a very carefully defined system with excited state populations between 0.5–5 % and with exchange rates of 10^2 – 10^3 s⁻¹.

Figure A.1. plots the relaxation rates determined in this study; these are also projected onto the crystal structure of α_1 -antitrypsin in Figure A.2. While some variation in relaxation rates is observed across the protein, given the asymmetric diffusion tensor expected given the α_1 -antitrypsin structure, a complete interpretation of these variations must necessarily wait for the model-free analysis to be completed using relaxation data obtained at multiple field strengths. Nevertheless, increased mobility is clearly discernible in loop regions (as increased R_1 , decreased R_2 and decreased NOE values), and in particular we observe that the reactive center loop exhibits large changes in its internal mobility across its sequence.

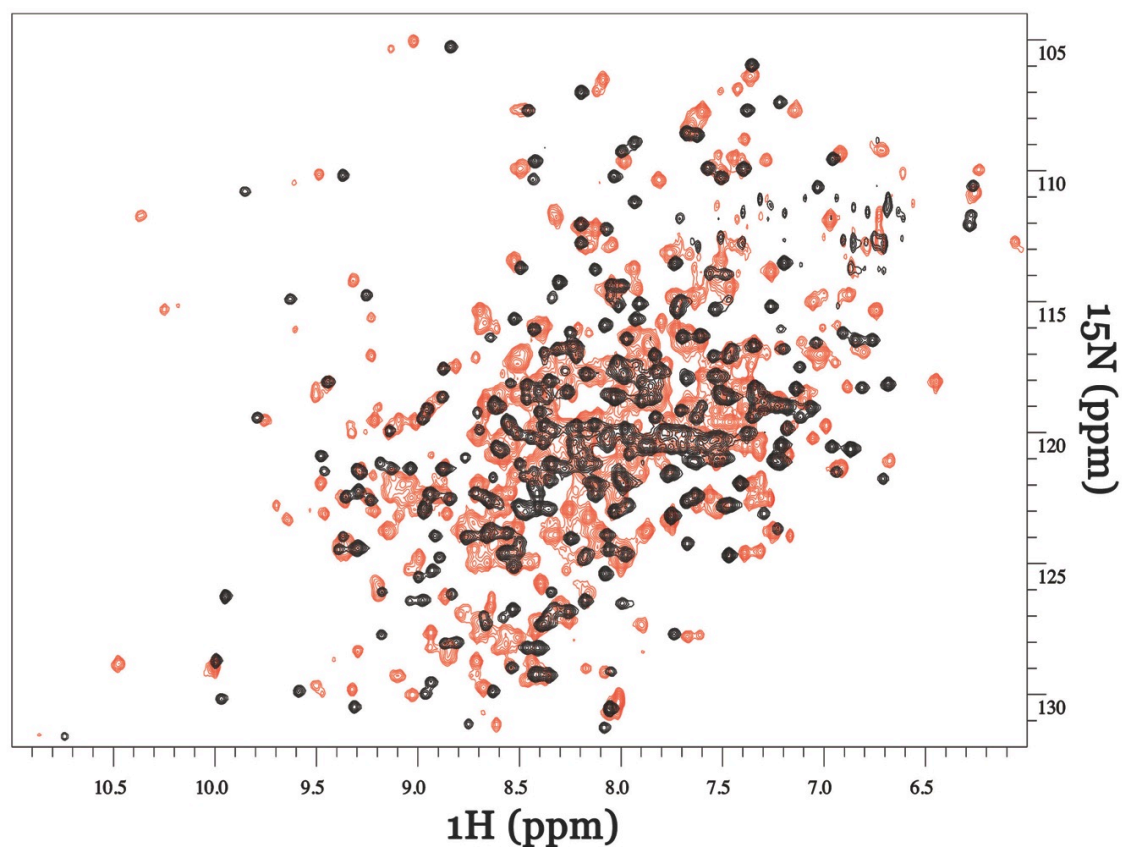


Appendix A Fig. 1: ^{15}N R_1 and R_2 relaxation rates, and $^{15}\text{N}\{-^1\text{H}\}$ NOE for α_1 -antitrypsin, recorded at 700 MHz, 298 K.

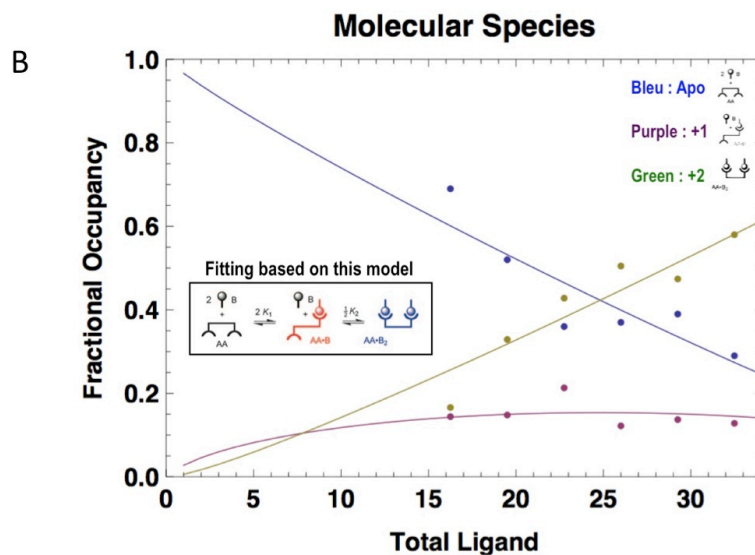
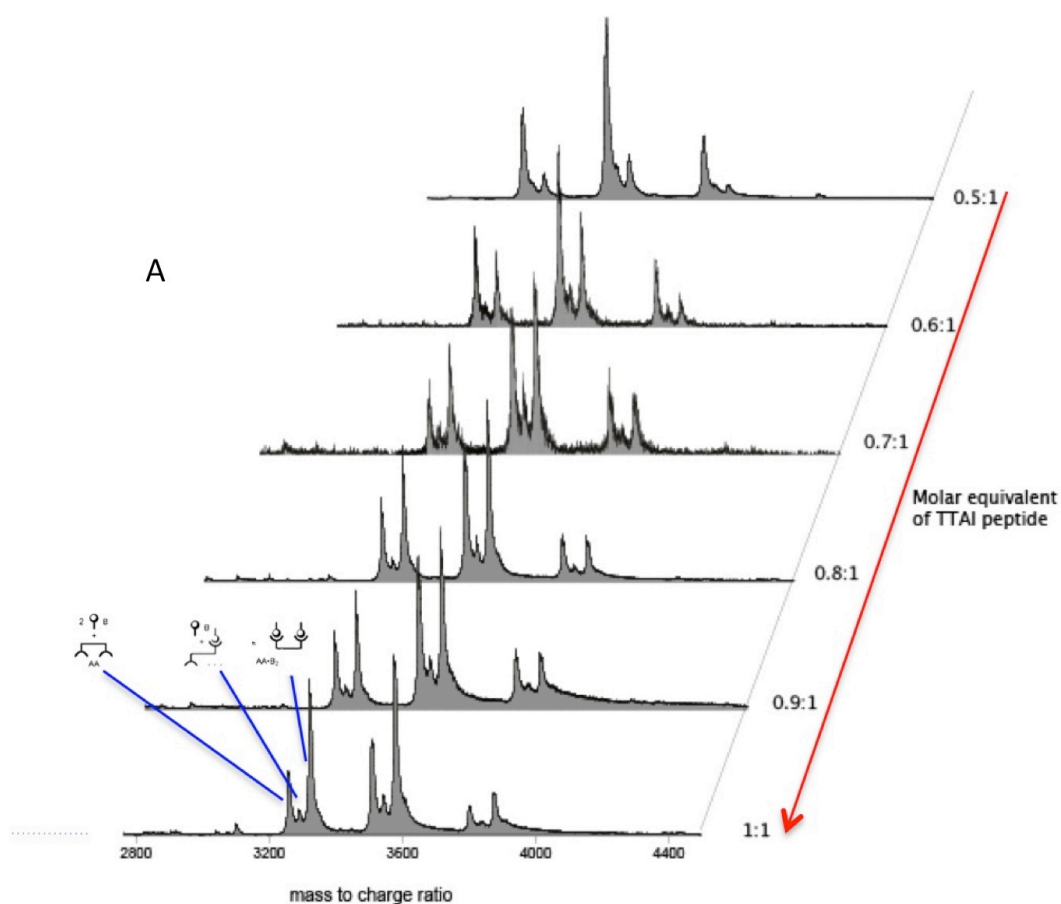


Appendix A Fig. 2 ^{15}N relaxation rates, R_1 , R_2 , and $^{15}\text{N}\{-^1\text{H}\}$ NOE depicted on the x-ray crystal structure of α_1 -antitrypsin (PDB: 3 NE4). Colours schemes are from blue to red, depicting from s^{-1} (R_1), a-b s^{-1} (R_2) and 0 - 1 ($^{15}\text{N}\{-^1\text{H}\}$ NOE).

Appendix B: Therapeutic strategies for blocking α_1 -Antitrypsin polymerisation: TTAI peptide binding



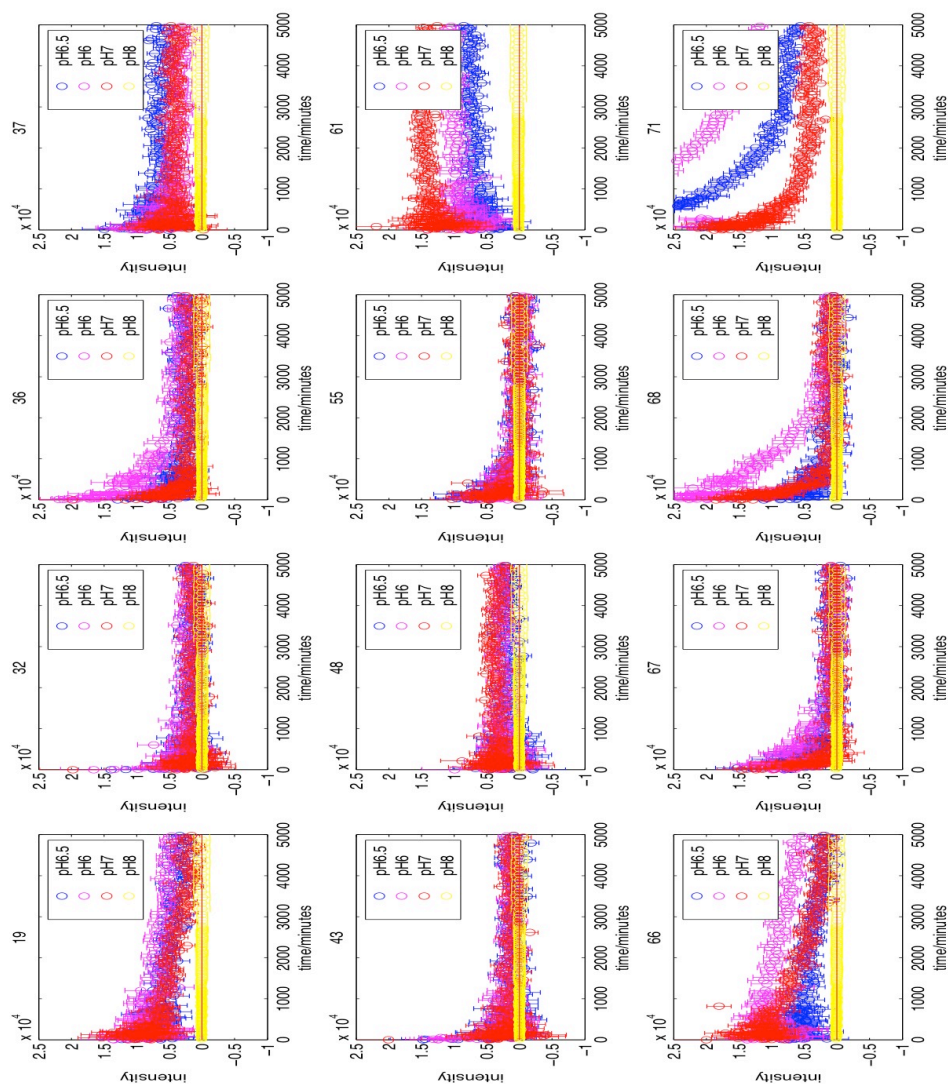
Appendix B Fig. 1. ^1H - ^{15}N HSQC spectra of α_1 -antitrypsin before (orange) and after (black) binding of the TTAI peptide (10:1 molar equivalence peptide:protein 37 °C) demonstrates some major conformational changes upon binding.

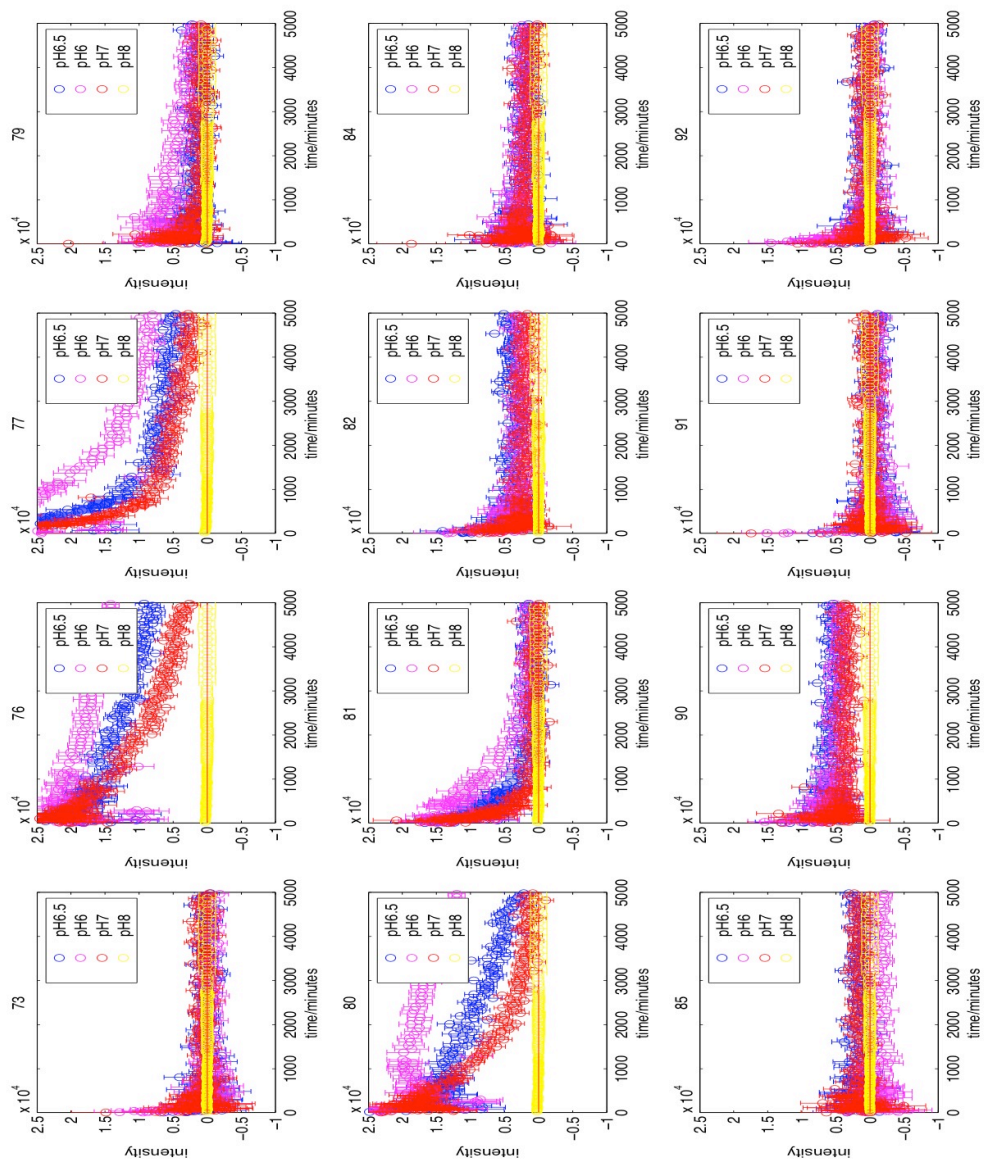


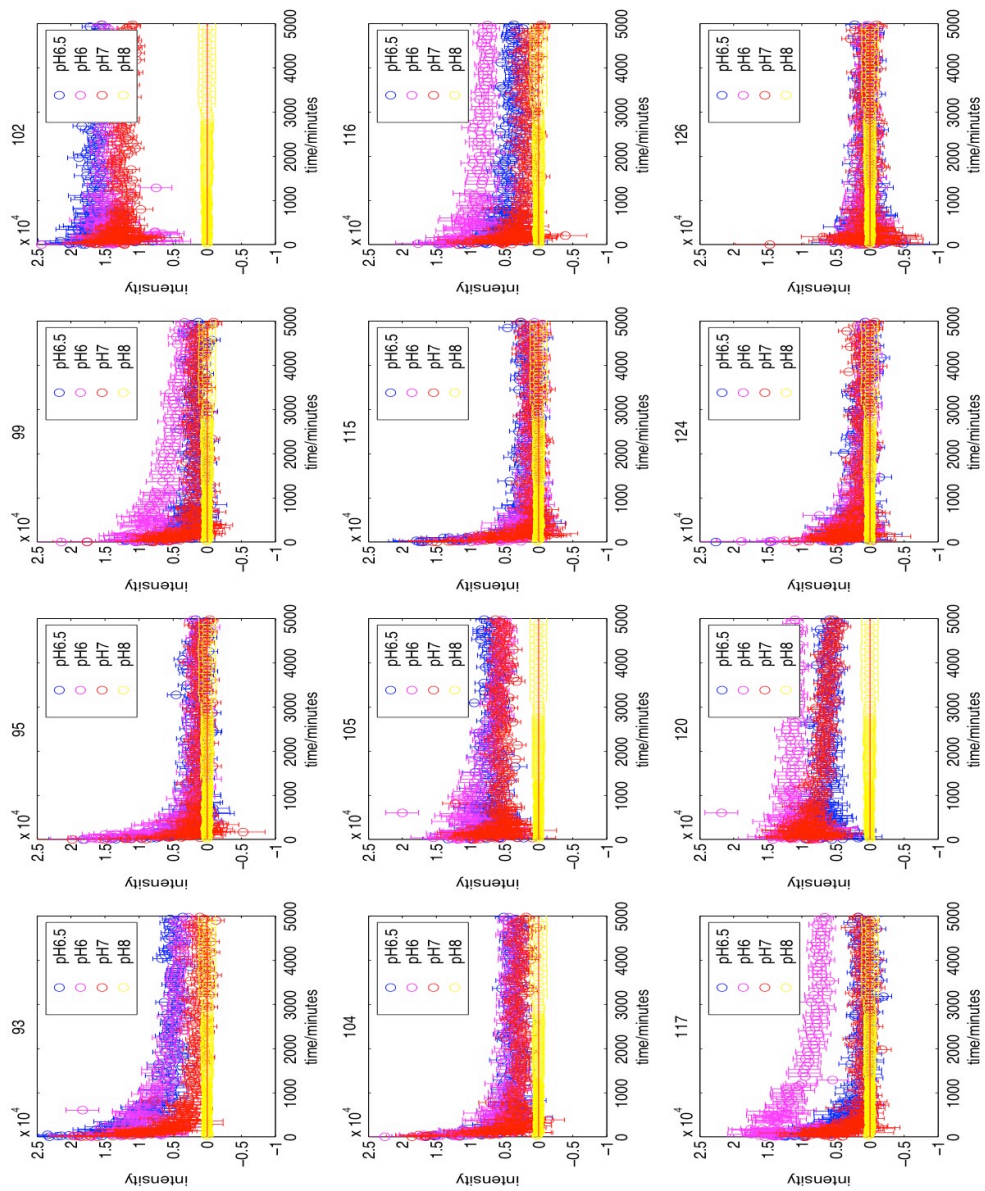
Appendix B Fig. 2. Ion Mobility Mass Spectrometry titration of α_1 -antitrypsin binding increasing concentrations of TTAI peptide. A- Stack plot of the +12, +13, +14 monomeric states signal obtained from 0.5:1 to 1:1 molar equivalent into 32.5 μ M of α_1 -antitrypsin. The charge states of α_1 -antitrypsin monomer appear as well-resolved triplet corresponding to the apo α_1 -antitrypsin, singly-bound α_1 -antitrypsin, and doubly-bound α_1 -antitrypsin. B- Illustrates fractional occupancy of α_1 -antitrypsin with increasing concentrations of ligand. Our data, suggests a cooperative binding, stoichiometry 2:1, for the binding of the TTAI peptide on α_1 -antitrypsin.

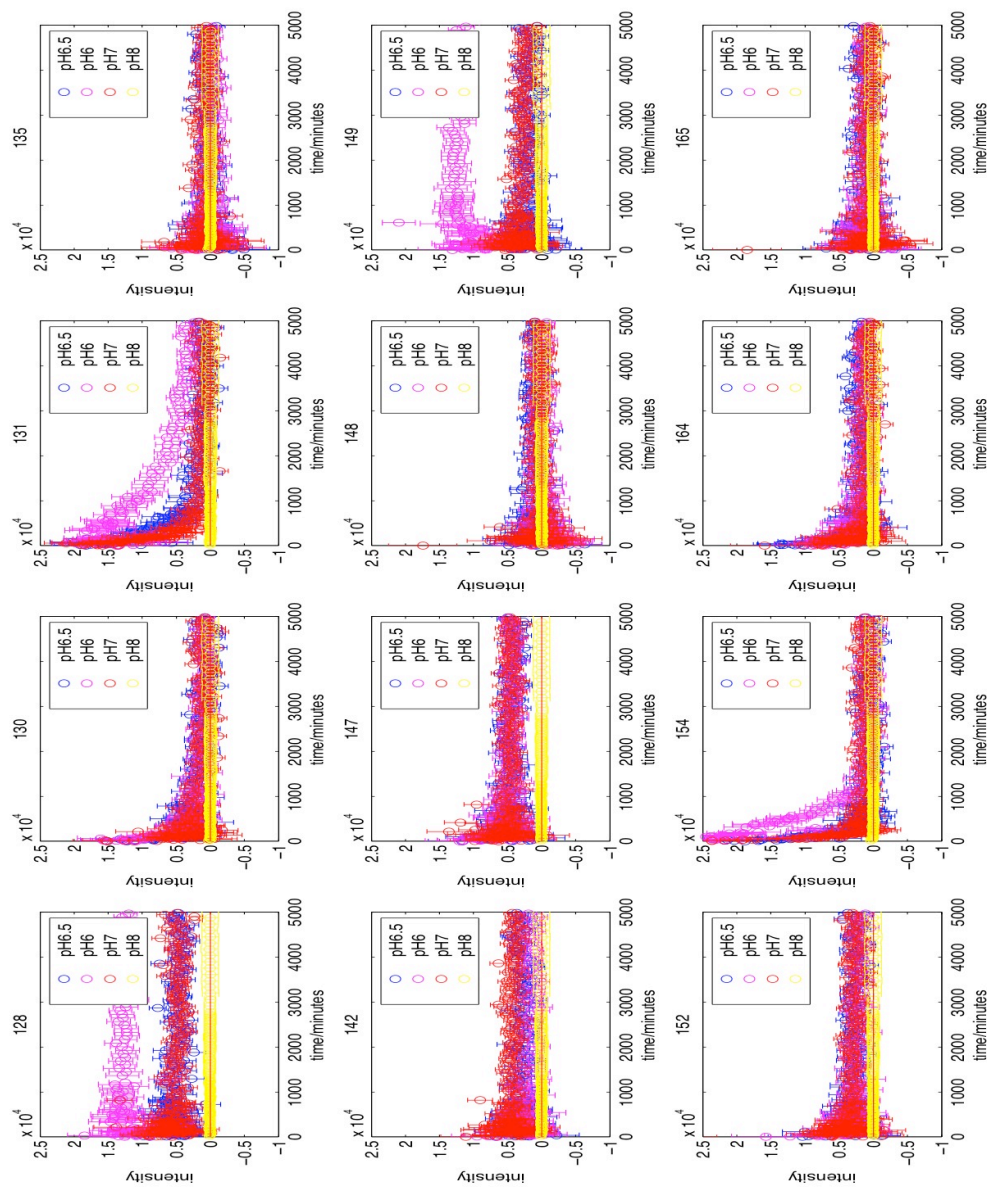
Appendix C: Fast Timescale Dynamics from HDXNMR experiments of α_1 -antitrypsin

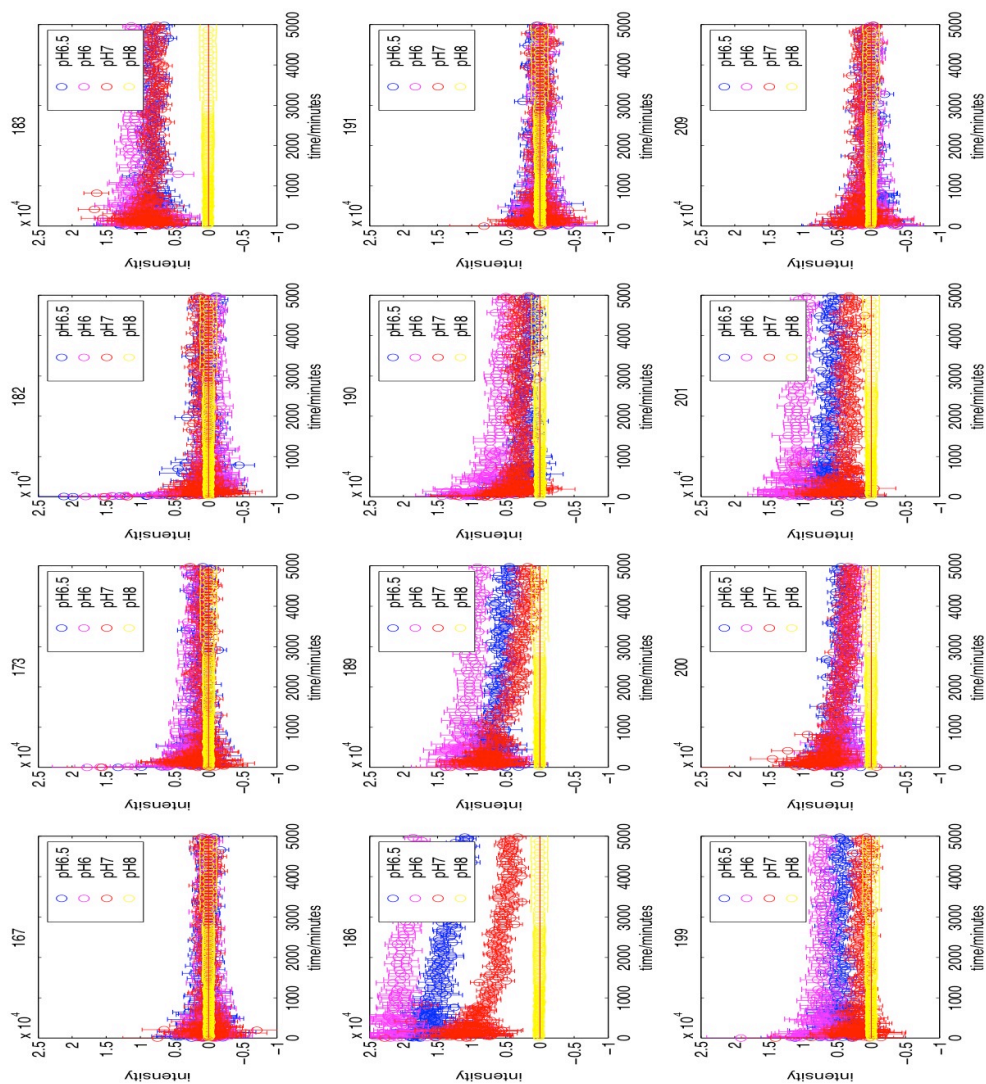
Plot of the raw intensity decay at a residue specific level for the HDXNMR analysis at pH 8.0, pH 7.0, pH 6.5, pH 6.0. Selected residues were subsequently selected for exponential fitting and free energy determination.

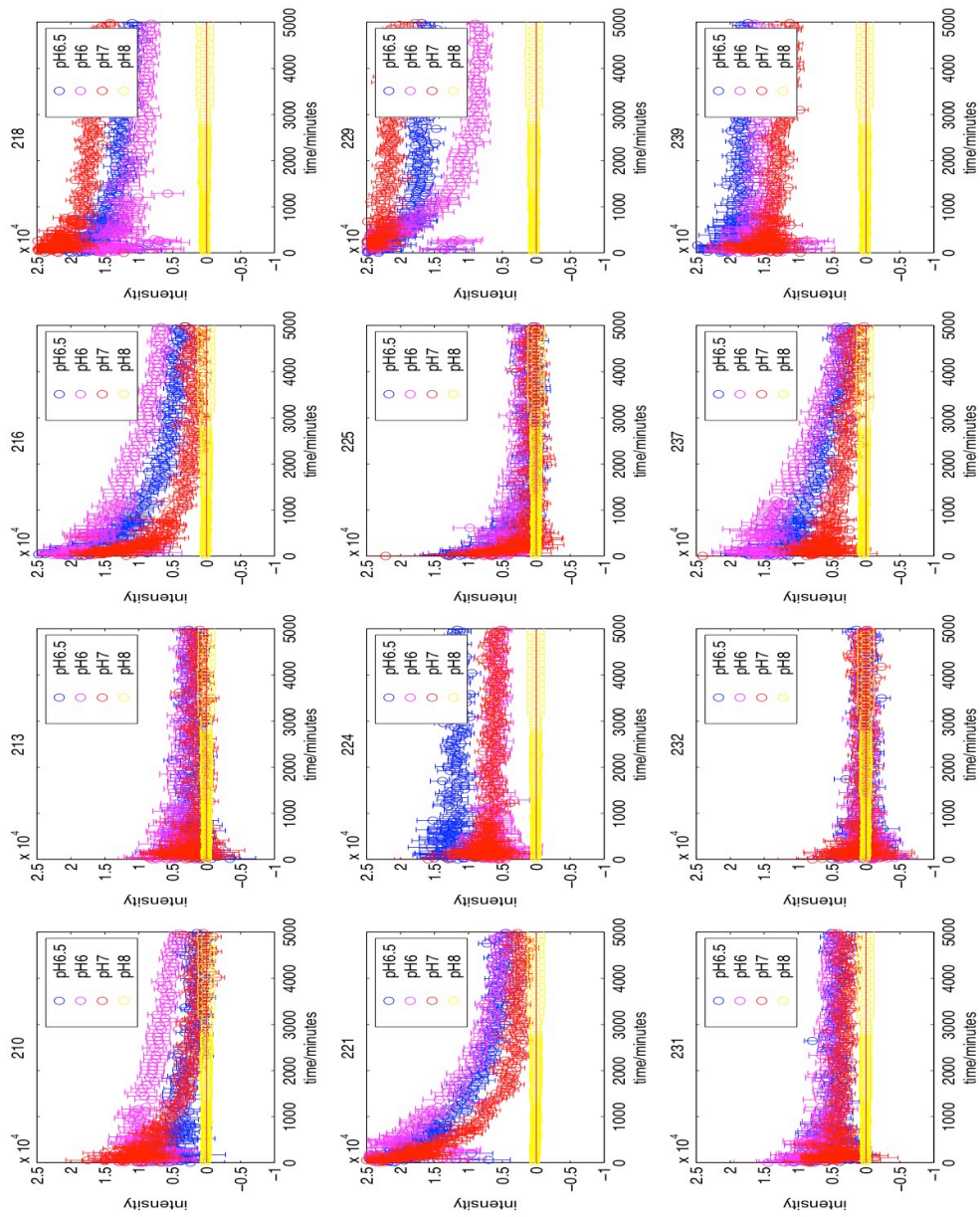


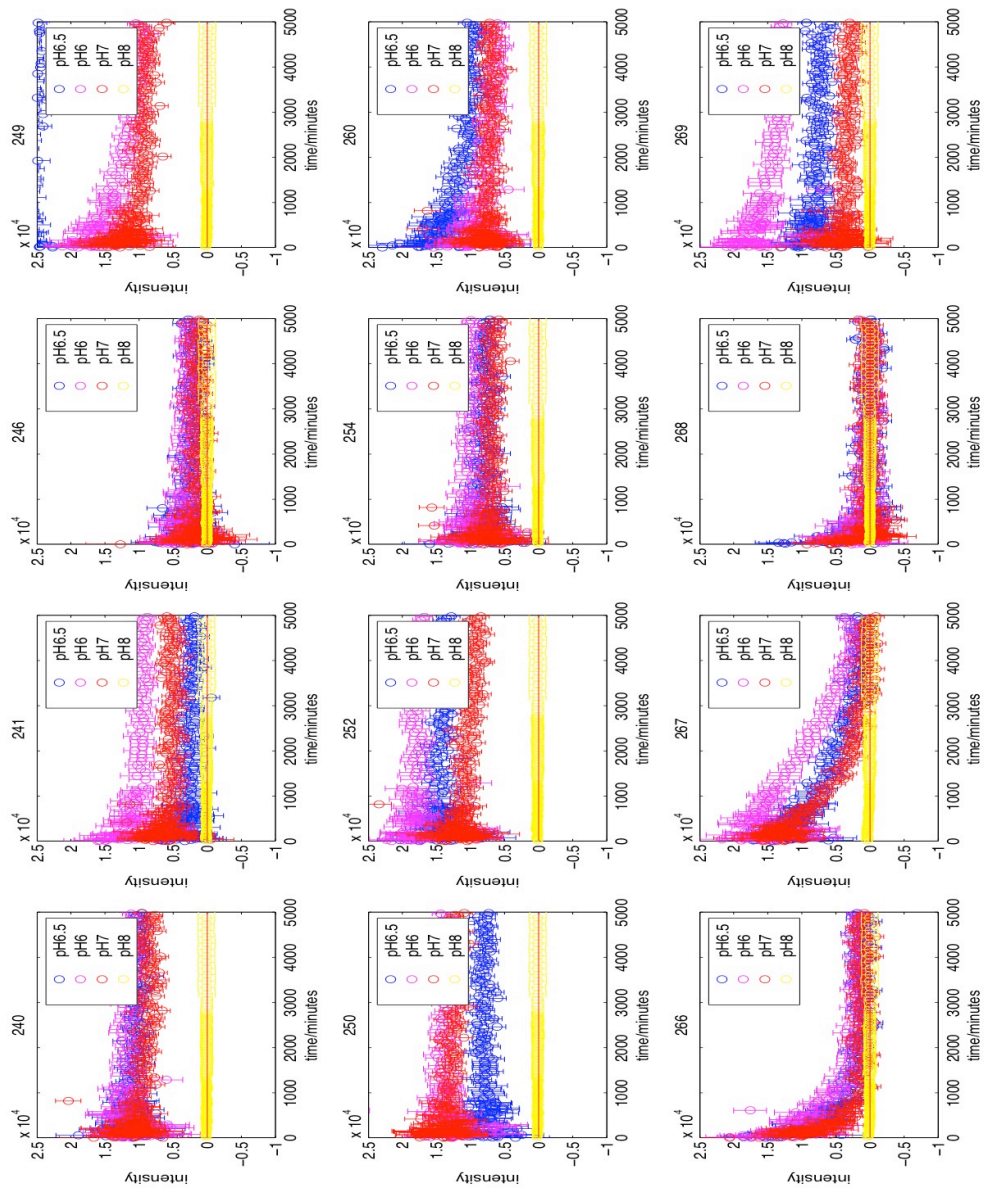


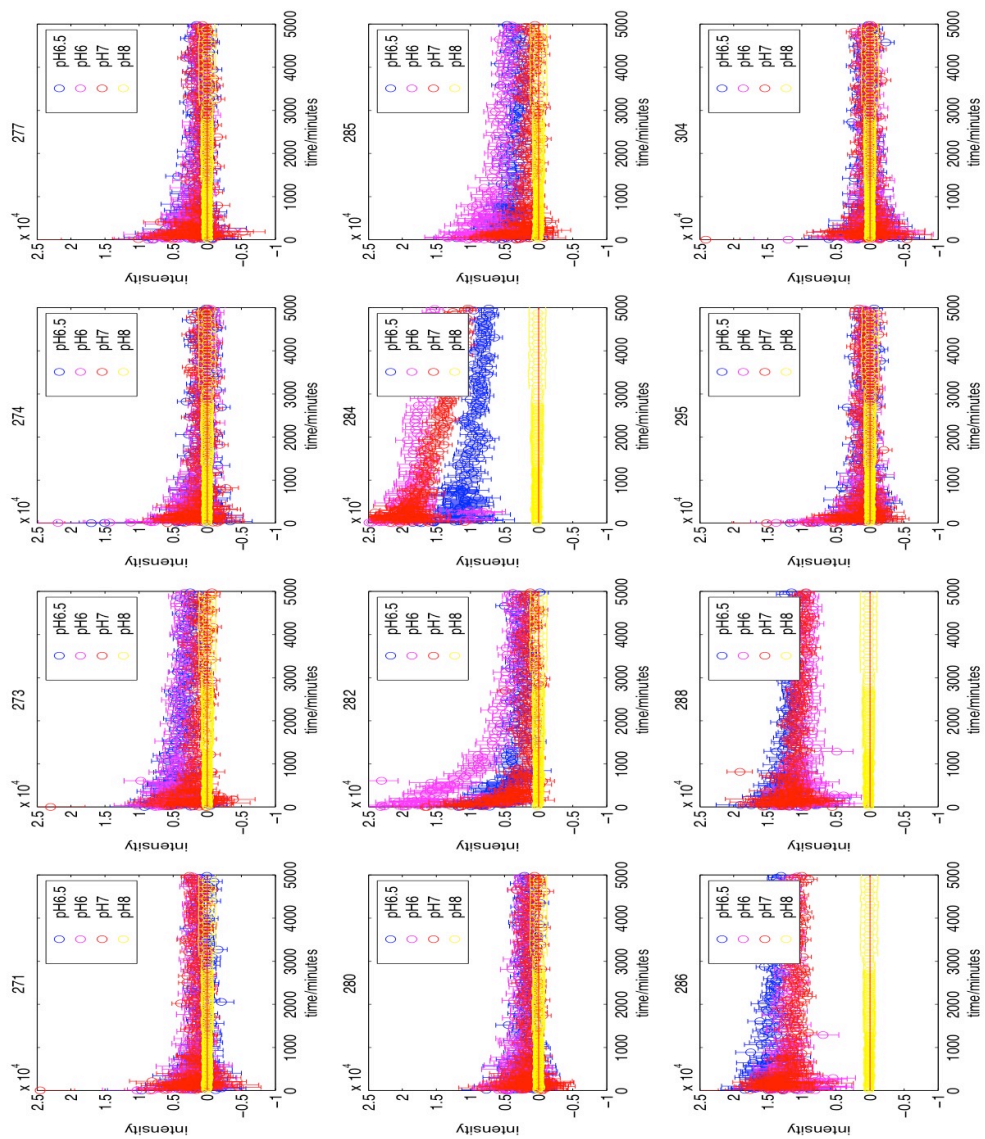


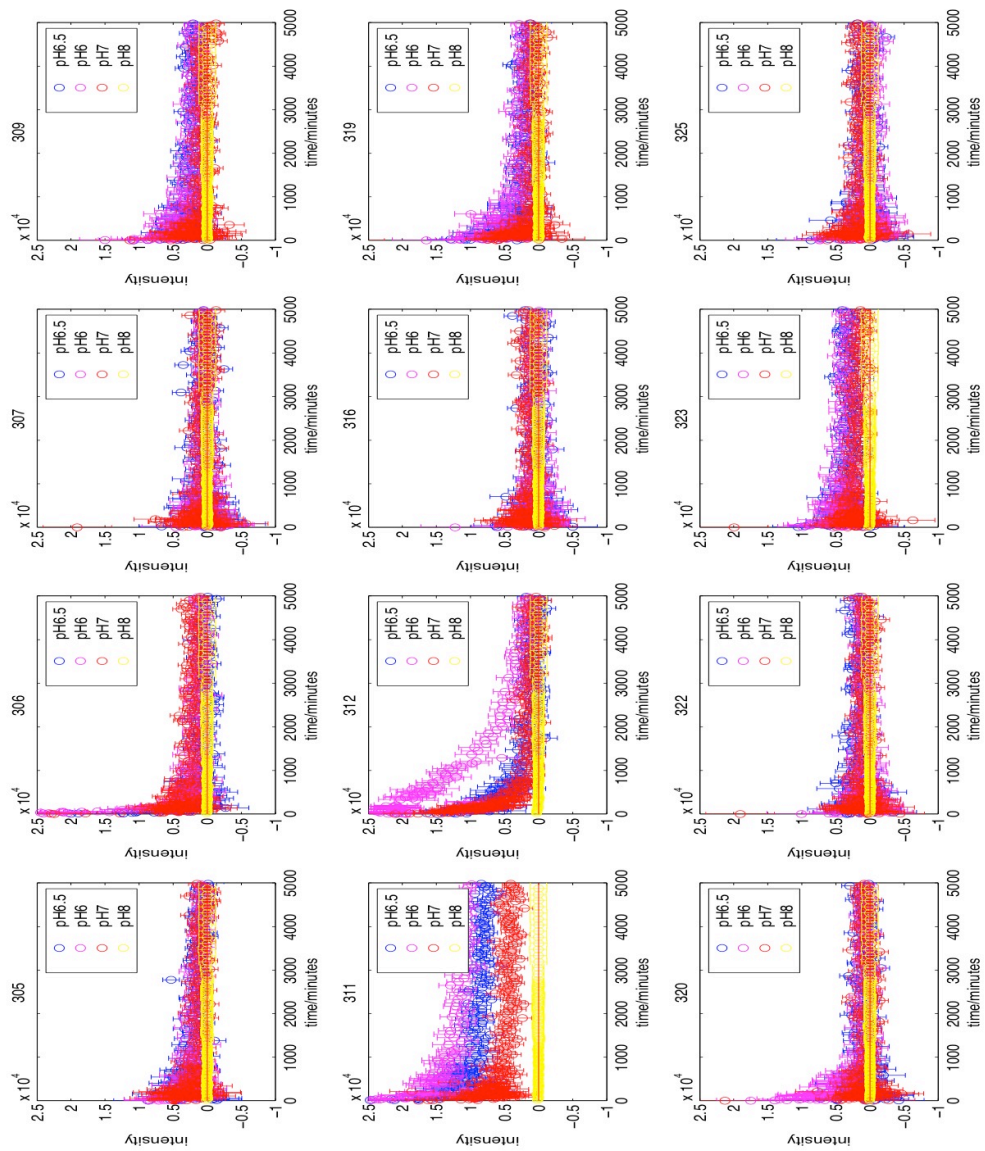


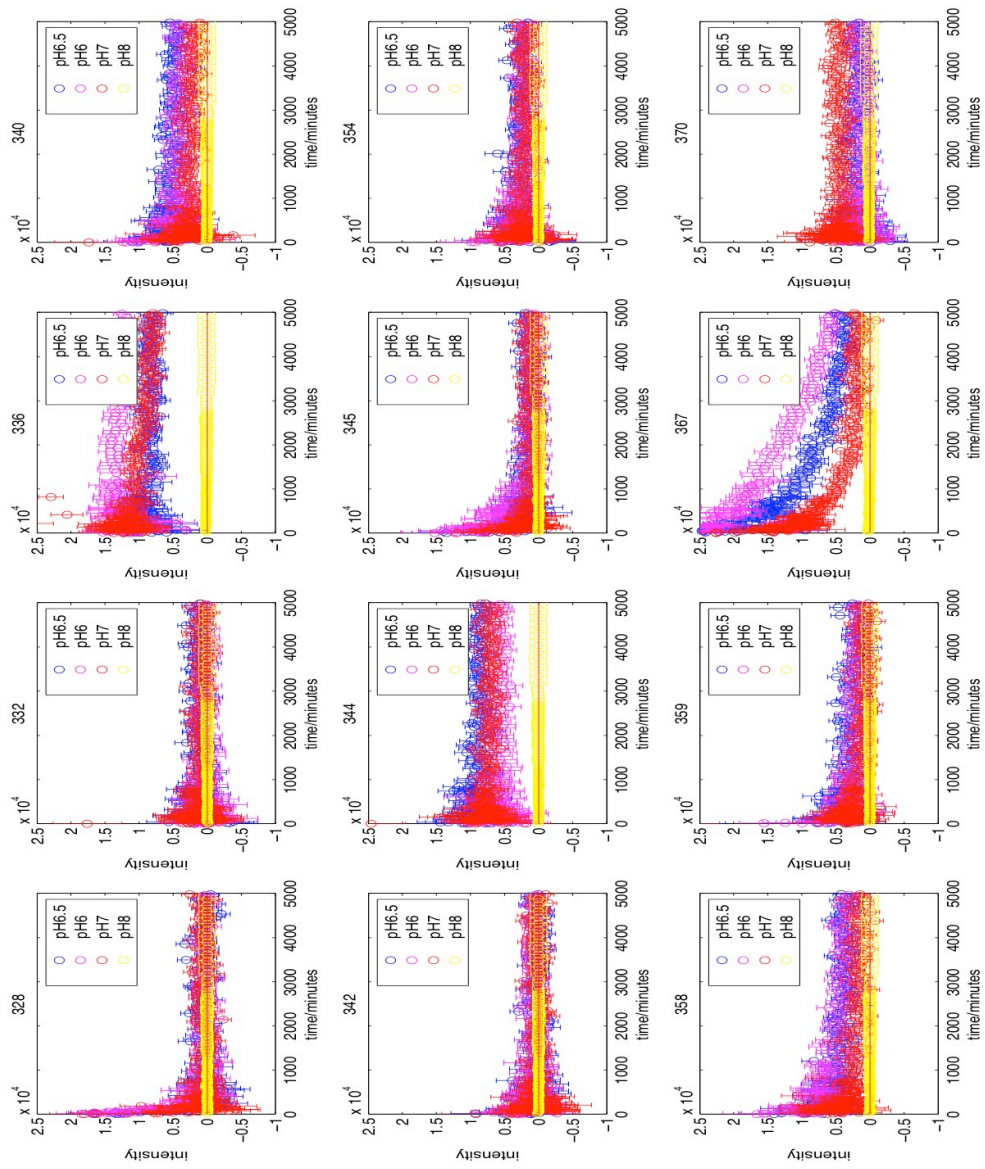


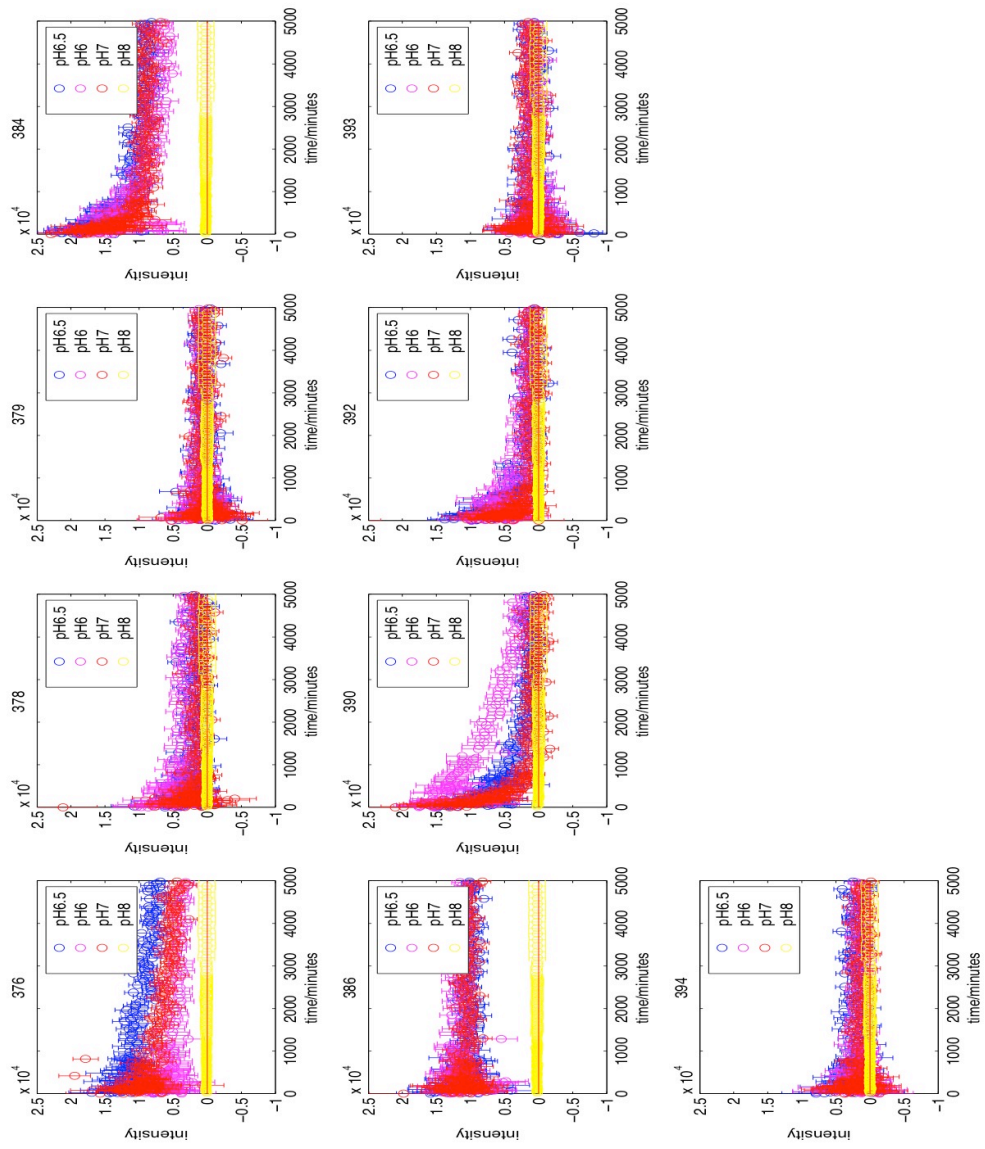












References

1. Copland, P. (2005). The book of life. *J Med Ethics* 31, 278-9.
2. Epstein, C. J. (1964). Relation of Protein Evolution to Tertiary Structure. *Nature* 203, 1350-2.
3. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223-30.
4. Anfinsen, B. C. (1972). The Nobel Prize in Chemistry 1972. In http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1972/press.html. The Royal Swedish Academy of Sciences
5. Isom, D. G., Castaneda, C. A., Cannon, B. R., Velu, P. D. & Garcia-Moreno, E. B (2010). Charges in the hydrophobic interior of proteins. *Proc Natl Acad Sci U S A* 107, 16096-100.
6. Zavodszky, A. S. J. K. S. O. L. B. P., Ed. (2007). Protein folding. Vol. 10. Berlin Heidelberg: Springer-Verlag
7. Makhatadze, G. I. & Privalov, P. L. (1995). Energetics of protein structure. *Adv Protein Chem* 47, 307-425.
8. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J* 10, 75-83.
9. Zwanzig, R., Szabo, A. & Bagchi, B. (1992). Levinthal's paradox. *Proc Natl Acad Sci U S A* 89, 20-2.
10. Sinnige, T., Karagöz, G Elif, and Rüdiger, Stefan GD. (Jun 2010). Protein Folding and Chaperones. eLS John Wiley & Sons Ltd, Chichester., <http://www.els.net> [doi: 10.1002/9780470015902.a0005721.pub2]

11. Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* 70, 697-701.
12. Harrison, S. C. & Durbin, R. (1985). Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci U S A* 82, 4028-30.
13. Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci U S A* 92, 10869-73.
14. Nolting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G. & Fersht, A. R. (1997). The folding pathway of a protein at high resolution from microseconds to seconds. *Proc Natl Acad Sci U S A* 94, 826-30.
15. Daggett, V. & Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem Sci* 28, 18-25.
16. Kuwajima, K., Nitta, K., Yoneyama, M. & Sugai, S. (1976). Three-state denaturation of α -lactalbumin by guanidine hydrochloride. *J Mol Biol* 106, 359-73.
17. Wong, K. P. & Tanford, C. (1973). Denaturation of bovine carbonic anhydrase B by guanidine hydrochloride. A process involving separable sequential conformational transitions. *J Biol Chem* 248, 8518-23.
18. Robson, B. & Pain, R. H. (1976). The mechanism of folding of globular proteins. Equilibria and kinetics of conformational transitions of penicillinase from *Staphylococcus aureus* involving a state of intermediate conformation. *Biochem J* 155, 331-44.
19. Nozaka, M., Kuwajima, K., Nitta, K. & Sugai, S. (1978). Detection and characterization of the intermediate on the folding pathway of human α -lactalbumin. *Biochemistry* 17, 3753-8.

20. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci* 24, 77-83.
21. Onuchic, J. N. & Wolynes, P. G. (2004). Theory of protein folding. *Curr Opin Struct Biol* 14, 70-5.
22. Jahn, T. R. & Radford, S. E. (2005). The Yin and Yang of protein folding. *FEBS J* 272, 5962-70.
23. Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett* 164, 21-4.
24. Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. (1986). Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: a comparative study of the folding reactions of α -lactalbumin and lysozyme. *Biochemistry* 25, 6965-72.
25. Ptitsyn, O. B. (1995). Molten globule and protein folding. *Adv Protein Chem* 47, 83-229.
26. Vendruscolo, M. & Dobson, C. M. (2005). Towards complete descriptions of the free-energy landscapes of proteins. *Philos Trans A Math Phys Eng Sci* 363, 433-50; discussion 450-2.
27. Chiti, F. & Dobson, C. M. (2009). Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5, 15-22.
28. Cabrita, L. D., Dobson, C. M. & Christodoulou, J. (2010). Protein folding on the ribosome. *Curr Opin Struct Biol* 20, 33-45.
29. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem Rev* 104, 3607-22.

30. Roder, H., Maki, K., Cheng, H. & Shastry, M. C. (2004). Rapid mixing methods for exploring the kinetics of protein folding. *Methods* 34, 15-27.
31. Chan, C. K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A. & Hofrichter, J. (1997). Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc Natl Acad Sci U S A* 94, 1779-84.
32. Kimura, T., Akiyama, S., Uzawa, T., Ishimori, K., Morishima, I., Fujisawa, T. & Takahashi, S. (2005). Specifically collapsed intermediate in the early stage of the folding of ribonuclease A. *J Mol Biol* 350, 349-62.
33. Roder, H. & Shastry, M. R. (1999). Methods for exploring early events in protein folding. *Curr Opin Struct Biol* 9, 620-6.
34. Williamson, J. R. (2009). The ribosome at atomic resolution. *Cell* 139, 1041-3.
35. Fedorov, A. N. & Baldwin, T. O. (1997). Cotranslational protein folding. *J Biol Chem* 272, 32715-8.
36. Cabrita, L. D., Hsu, S. T., Launay, H., Dobson, C. M. & Christodoulou, J. (2009). Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc Natl Acad Sci U S A* 106, 22239-44.
37. Christodoulou, J., Larsson, G., Fucini, P., Connell, S. R., Pertinhez, T. A., Hanson, C. L., Redfield, C., Nierhaus, K. H., Robinson, C. V., Schleucher, J. & Dobson, C. M. (2004). Heteronuclear NMR investigations of dynamic regions of intact *Escherichia coli* ribosomes. *Proc Natl Acad Sci U S A* 101, 10949-54.
38. Hsu, S. T., Cabrita, L. D., Fucini, P., Christodoulou, J. & Dobson, C. M. (2009). Probing side-chain dynamics of a ribosome-bound nascent chain using methyl NMR spectroscopy. *J Am Chem Soc* 131, 8366-7.

39. Hsu, S. T., Fucini, P., Cabrita, L. D., Launay, H., Dobson, C. M. & Christodoulou, J. (2007). Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc Natl Acad Sci U S A* 104, 16516-21.
40. Steitz, T. A. (2008). A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* 9, 242-53.
41. Bashan, A. & Yonath, A. (2008). Correlating ribosome function with high-resolution structures. *Trends Microbiol* 16, 326-35.
42. Mitra, K. & Frank, J. (2006). Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annu Rev Biophys Biomol Struct* 35, 299-317.
43. Steitz, J. A. (1969). Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature* 224, 957-64.
44. Schaffitzel, C. & Ban, N. (2007). Generation of ribosome nascent chain complexes for structural and functional studies. *J Struct Biol* 158, 463-71.
45. Carrell, R. W. & Lomas, D. A. (2002). α_1 -antitrypsin deficiency--a model for conformational diseases. *N Engl J Med* 346, 45-53.
46. Lomas, D. A. & Carrell, R. W. (2002). Serpinopathies and the conformational dementias. *Nat Rev Genet* 3, 759-68.
47. Dobson, C. M. (2003). Protein folding and misfolding. *Nature* 426, 884-90.
48. Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem Sci* 24, 329-32.

49. Amaral, M. D. (2004). CFTR and chaperones: processing and degradation. *J Mol Neurosci* 23, 41-8.
50. Van Goor, F., Hadida, S., Grootenhuys, P. D., Burton, B., Cao, D., Neuberger, T., Turnbull, A., Singh, A., Joubran, J., Hazlewood, A., Zhou, J., McCartney, J., Arumugam, V., Decker, C., Yang, J., Young, C., Olson, E. R., Wine, J. J., Frizzell, R. A., Ashlock, M. & Negulescu, P. (2009). Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc Natl Acad Sci U S A* 106, 18825-30.
51. Accurso, F. J., Rowe, S. M., Clancy, J. P., Boyle, M. P., Dunitz, J. M., Durie, P. R., Sagel, S. D., Hornick, D. B., Konstan, M. W., Donaldson, S. H., Moss, R. B., Pilewski, J. M., Rubenstein, R. C., Uluer, A. Z., Aitken, M. L., Freedman, S. D., Rose, L. M., Mayer-Hamblett, N., Dong, Q., Zha, J., Stone, A. J., Olson, E. R., Ordonez, C. L., Campbell, P. W., Ashlock, M. A. & Ramsey, B. W. (2010). Effect of VX-770 in persons with cystic fibrosis and the G551D-CFTR mutation. *N Engl J Med* 363, 1991-2003.
52. Dobson, C. M. (2004). Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* 15, 3-16.
53. Dobson, C. M. (2001). Protein folding and its links with human disease. *Biochem Soc Symp*, 1-26.
54. Gooptu, B. & Lomas, D. A. (2009). Conformational pathology of the serpins: themes, variations, and therapeutic strategies. *Annu Rev Biochem* 78, 147-76.
55. Hirsch, E. C. (1994). Biochemistry of Parkinson's disease with special reference to the dopaminergic systems. *Mol Neurobiol* 9, 135-42.

56. Stefani, M. & Dobson, C. M. (2003). Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* 81, 678-99.
57. Forloni, G., Terreni, L., Bertani, I., Fogliarino, S., Invernizzi, R., Assini, A., Ribizzi, G., Negro, A., Calabrese, E., Volonte, M. A., Mariani, C., Franceschi, M., Tabaton, M. & Bertoli, A. (2002). Protein misfolding in Alzheimer's and Parkinson's disease: genetics and molecular mechanisms. *Neurobiol Aging* 23, 957-76.
58. Rubinsztein, D. C. & Carmichael, J. (2003). Huntington's disease: molecular basis of neurodegeneration. *Expert Rev Mol Med* 5, 1-21.
59. Weissmann, C. (1991). A 'unified theory' of prion propagation. *Nature* 352, 679-83.
60. Carrell, R. W., Whisstock, J. & Lomas, D. A. (1994). Conformational changes in serpins and the mechanism of α_1 -antitrypsin deficiency. *Am J Respir Crit Care Med* 150, S171-5.
61. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223-30.
62. Perlmutter, D. H. (2009). Autophagic disposal of the aggregation-prone protein that causes liver inflammation and carcinogenesis in α_1 -antitrypsin deficiency. *Cell Death Differ* 16, 39-45.
63. Perlmutter, D. H. (2006). The role of autophagy in α_1 -antitrypsin deficiency: a specific cellular response in genetic diseases associated with aggregation-prone proteins. *Autophagy* 2, 258-63.
64. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27, 157-62.

65. Huntington, J. A., Read, R. J. & Carrell, R. W. (2000). Structure of a serpin-protease complex shows inhibition by deformation. *Nature* 407, 923-6.
66. Dolmer, K. & Gettins, P. G. (2012). How the serpin α 1-proteinase inhibitor folds. *J Biol Chem* 287, 12425-32.
67. Whisstock, J. C. & Bottomley, S. P. (2006). Molecular gymnastics: serpin structure, folding and misfolding. *Curr Opin Struct Biol* 16, 761-8.
68. Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984). Human α 1-proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J Mol Biol* 177, 531-57.
69. Carrell, R. W. (1986). α 1-antitrypsin: molecular pathology, leukocytes, and tissue damage. *J Clin Invest* 78, 1427-31.
70. Mottonen, J., Strand, A., Symersky, J., Sweet, R. M., Danley, D. E., Geoghegan, K. F., Gerard, R. D. & Goldsmith, E. J. (1992). Structural basis of latency in plasminogen activator inhibitor-1. *Nature* 355, 270-3.
71. Wardell, M. R., Chang, W. S., Bruce, D., Skinner, R., Lesk, A. M. & Carrell, R. W. (1997). Preparative induction and characterization of L-antithrombin: a structural homologue of latent plasminogen activator inhibitor-1. *Biochemistry* 36, 13133-42.
72. Lomas, D. A., Elliott, P. R., Chang, W. S., Wardell, M. R. & Carrell, R. W. (1995). Preparation and characterization of latent α 1-antitrypsin. *J Biol Chem* 270, 5282-8.

73. Koloczek, H., Banbula, A., Salvesen, G. S. & Potempa, J. (1996). Serpin α 1proteinase inhibitor probed by intrinsic tryptophan fluorescence spectroscopy. *Protein Sci* 5, 2226-35.
74. Bottomley, S. P., Hopkins, P. C. & Whisstock, J. C. (1998). α 1-antitrypsin polymerisation can occur by both loop A and C sheet mechanisms. *Biochem Biophys Res Commun* 251, 1-5.
75. Onda, M., Belorgey, D., Sharp, L. K. & Lomas, D. A. (2005). Latent S49P neuroserpin forms polymers in the dementia familial encephalopathy with neuroserpin inclusion bodies. *J Biol Chem* 280, 13735-41.
76. Chang, W. S. & Lomas, D. A. (1998). Latent α 1-antichymotrypsin. A molecular explanation for the inactivation of α 1-antichymotrypsin in chronic bronchitis and emphysema. *J Biol Chem* 273, 3695-701.
77. Gooptu, B., Hazes, B., Chang, W. S., Dafforn, T. R., Carrell, R. W., Read, R. J. & Lomas, D. A. (2000). Inactive conformation of the serpin α 1-antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease. *Proc Natl Acad Sci U S A* 97, 67-72.
78. Schulze, A. J., Frohnert, P. W., Engh, R. A. & Huber, R. (1992). Evidence for the extent of insertion of the active site loop of intact α 1 proteinase inhibitor in beta-sheet A. *Biochemistry* 31, 7560-5.
79. Bjork, I., Ylinenjarvi, K., Olson, S. T. & Bock, P. E. (1992). Conversion of antithrombin from an inhibitor of thrombin to a substrate with reduced heparin affinity and enhanced conformational stability by binding of a tetradecapeptide corresponding to the P1 to P14 region of the putative reactive bond loop of the inhibitor. *J Biol Chem* 267, 1976-82.

80. Schulze, A. J., Baumann, U., Knof, S., Jaeger, E., Huber, R. & Laurell, C. B. (1990). Structural transition of α_1 -antitrypsin by a peptide sequentially similar to beta-strand s4A. *Eur J Biochem* 194, 51-6.
81. Tew, D. J. & Bottomley, S. P. (2001). Probing the equilibrium denaturation of the serpin α_1 -antitrypsin with single tryptophan mutants; evidence for structure in the urea unfolded state. *J Mol Biol* 313, 1161-9.
82. Pearce, M. C., Rubin, H. & Bottomley, S. P. (2000). Conformational change and intermediates in the unfolding of α_1 -antichymotrypsin. *J Biol Chem* 275, 28513-8.
83. James, E. L., Whisstock, J. C., Gore, M. G. & Bottomley, S. P. (1999). Probing the unfolding pathway of α_1 -antitrypsin. *J Biol Chem* 274, 9482-8.
84. Benning, L. N., Whisstock, J. C., Sun, J., Bird, P. I. & Bottomley, S. P. (2004). The human serpin proteinase inhibitor-9 self-associates at physiological temperatures. *Protein Sci* 13, 1859-64.
85. Cabrita, L. D., Irving, J. A., Pearce, M. C., Whisstock, J. C. & Bottomley, S. P. (2007). Aeropin from the extremophile *Pyrobaculum aerophilum* bypasses the serpin misfolding trap. *J Biol Chem* 282, 26802-9.
86. Zhang, Q., Buckle, A. M., Law, R. H., Pearce, M. C., Cabrita, L. D., Lloyd, G. J., Irving, J. A., Smith, A. I., Ruzyla, K., Rossjohn, J., Bottomley, S. P. & Whisstock, J. C. (2007). The N terminus of the serpin, tengpin, functions to trap the metastable native state. *EMBO Rep* 8, 658-63.
87. Powell, L. M. & Pain, R. H. (1992). Effects of glycosylation on the folding and stability of human, recombinant and cleaved α_1 -antitrypsin. *J Mol Biol* 224, 241-52.

88. Refaei, M. A., Combs, A., Kojetin, D. J., Cavanagh, J., Caperelli, C., Rance, M., Sapiro, J. & Tsang, P. (2011). Observing selected domains in multi-domain proteins via sortase-mediated ligation and NMR spectroscopy. *J Biomol NMR* 49, 3-7.
89. Herve, M. & Ghelis, C. (1990). Conformational changes in intact and papain-modified α 1-proteinase inhibitor induced by guanidinium chloride. *Eur J Biochem* 191, 653-8.
90. Tsutsui, Y., Kuri, B., Sengupta, T. & Wintrode, P. L. (2008). The structural basis of serpin polymerization studied by hydrogen/deuterium exchange and mass spectrometry. *J Biol Chem* 283, 30804-11.
91. Tsutsui, Y. & Wintrode, P. L. (2007). Hydrogen/deuterium exchange-mass spectrometry: a powerful tool for probing protein structure, dynamics and interactions. *Curr Med Chem* 14, 2344-58.
92. Tsutsui, Y., Liu, L., Gershenson, A. & Wintrode, P. L. (2006). The conformational dynamics of a metastable serpin studied by hydrogen exchange and mass spectrometry. *Biochemistry* 45, 6561-9.
93. Ekeowa, U. I., Freeke, J., Miranda, E., Gooptu, B., Bush, M. F., Perez, J., Teckman, J., Robinson, C. V. & Lomas, D. A. (2010). Defining the mechanism of polymerization in the serpinopathies. *Proc Natl Acad Sci U S A* 107, 17146-51.
94. Tsutsui, Y., Dela Cruz, R. & Wintrode, P. L. (2012). Folding mechanism of the metastable serpin α 1-antitrypsin. *Proc Natl Acad Sci U S A* 109, 4467-72.

95. Lomas, D. A., Evans, D. L., Finch, J. T. & Carrell, R. W. (1992). The mechanism of Z α_1 -antitrypsin accumulation in the liver. *Nature* 357, 605-7.
96. Gooptu, B. & Lomas, D. A. (2008). Polymers and inflammation: disease mechanisms of the serpinopathies. *J Exp Med* 205, 1529-34.
97. Yamasaki, M., Li, W., Johnson, D. J. & Huntington, J. A. (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* 455, 1255-8.
98. Yamasaki, M., Sendall, T. J., Harris, L. E., Lewis, G. M. & Huntington, J. A. (2010). Loop-sheet mechanism of serpin polymerization tested by reactive center loop mutations. *J Biol Chem* 285, 30752-8.
99. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C. & Huntington, J. A. (2011). Molecular basis of α_1 -antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep* 12, 1011-7.
100. Bottomley, S. P. (2010). The folding pathway of α_1 -antitrypsin: avoiding the unavoidable. *Proc Am Thorac Soc* 7, 404-7.
101. Dafforn, T. R., Pike, R. N. & Bottomley, S. P. (2004). Physical characterization of serpin conformations. *Methods* 32, 150-8.
102. Lomas, D. A., Belorgey, D., Mallya, M., Onda, M., Kinghorn, K. J., Sharp, L. K., Phillips, R. L., Page, R., Crowther, D. C. & Miranda, E. (2004). Polymerisation underlies α_1 -antitrypsin deficiency, dementia and other serpinopathies. *Front Biosci* 9, 2873-91.
103. Dafforn, T. R., Mahadeva, R., Elliott, P. R., Sivasothy, P. & Lomas, D. A. (1999). A kinetic mechanism for the polymerization of α_1 -antitrypsin. *J Biol Chem* 274, 9548-55.

104. Mast, A. E., Enghild, J. J. & Salvesen, G. (1992). Conformation of the reactive site loop of α_1 -proteinase inhibitor probed by limited proteolysis. *Biochemistry* 31, 2720-8.
105. Huntington, J. A., Pannu, N. S., Hazes, B., Read, R. J., Lomas, D. A. & Carrell, R. W. (1999). A 2.6 Å structure of a serpin polymer and implications for conformational disease. *J Mol Biol* 293, 449-55.
106. Dunstone, M. A., Dai, W., Whisstock, J. C., Rossjohn, J., Pike, R. N., Feil, S. C., Le Bonniec, B. F., Parker, M. W. & Bottomley, S. P. (2000). Cleaved antitrypsin polymers at atomic resolution. *Protein Sci* 9, 417-20.
107. Fitton, H. L., Pike, R. N., Carrell, R. W. & Chang, W. S. (1997). Mechanisms of antithrombin polymerisation and heparin activation probed by the insertion of synthetic reactive loop peptides. *Biol Chem* 378, 1059-63.
108. Elliott, P. R., Lomas, D. A., Carrell, R. W. & Abrahams, J.-P. (1996). Inhibitory conformation of the reactive loop of α_1 -antitrypsin. *Nat. Struct. Biol.* 3, 676-681.
109. Dunstone, M. A., Dai, W., Whisstock, J. C., Rossjohn, J., Pike, R. N., Feil, S. C., Le Bonniec, B. F., Parker, M. W. & Bottomley, S. P. (2000). Cleaved antitrypsin polymers at atomic resolution. *Protein Sci.* 9, 417-20.
110. Huntington, J. A., Pannu, N. S., Hazes, B., Read, R., Lomas, D. A. & Carrell, R. W. (1999). A 2.6Å structure of a serpin polymer and implications for conformational disease. *J. Mol. Biol.* 293, 449-455.
111. Gooptu, B., Hazes, B., Chang, W.-S. W., Dafforn, T. R., Carrell, R. W., Read, R. & Lomas, D. A. (2000). Inactive conformation of the serpin α_1 -antichymotrypsin indicates two stage insertion of the reactive

- loop; implications for inhibitory function and conformational disease. *Proc. Natl. Acad. Sci (USA)* 97, 67-72.
112. Im, H., Woo, M.-S., Hwang, K. Y. & Yu, M.-H. (2002). Interactions causing the kinetic trap in serpin protein folding. *J. Biol. Chem.* 277, 46347-46354.
 113. Haq, I., Irving, J. A., Faull, S. V., Dickens, J. A., Ordonez, A., Belorgey, D., Gooptu, B. & Lomas, D. A. (2013). Reactive centre loop mutants of α_1 -antitrypsin reveal position-specific effects on intermediate formation along the polymerization pathway. *Biosci Rep* 33.
 114. Yamasaki, M., Li, W., Johnson, D. J. & Huntington, J. A. (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* 455, 1255-1258.
 115. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C. & Huntington, J. A. (2011). Molecular basis of α_1 -antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep* 12, 1011-7.
 116. Rabi, II, Zacharias, J. R., Millman, S. & Kusch, P. (1992). Milestones in magnetic resonance: 'a new method of measuring nuclear magnetic moment' . 1938. *J Magn Reson Imaging* 2, 131-3.
 117. Wuthrich, K. (1989). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* 243, 45-50.
 118. Wuthrich, K. (1989). Determination of three-dimensional protein structures in solution by nuclear magnetic resonance: an overview. *Methods Enzymol* 177, 125-31.
 119. Dobson, C. M. (1991). NMR spectroscopy and protein folding: studies of lysozyme and α -lactalbumin. *Ciba Found Symp* 161, 167-81; discussion 181-9.

120. Englander, S. W. & Mayne, L. (1992). Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu Rev Biophys Biomol Struct* 21, 243-65.
121. Hansen, D. F., Yang, D., Feng, H., Zhou, Z., Wiesner, S., Bai, Y. & Kay, L. E. (2007). An exchange-free measure of ¹⁵N transverse relaxation: an NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. *J Am Chem Soc* 129, 11468-79.
122. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430, 586-90.
123. Dobson, C. M. & Hore, P. J. (1998). Kinetic studies of protein folding using NMR spectroscopy. *Nat Struct Biol* 5 Suppl, 504-7.
124. Neudecker, P., Lundstrom, P. & Kay, L. E. (2009). Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys J* 96, 2045-54.
125. Vendruscolo, M., Paci, E., Karplus, M. & Dobson, C. M. (2003). Structures and relative free energies of partially folded states of proteins. *Proc Natl Acad Sci U S A* 100, 14817-21.
126. Vendruscolo, M., Zurdo, J., MacPhee, C. E. & Dobson, C. M. (2003). Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Philos Trans A Math Phys Eng Sci* 361, 1205-22.
127. Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D. & Alber, T. (2009). Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462, 669-73.

128. Ivano Bertini, K. S. M., Giacomo Parigi. (2012). NMR of Biomolecules: Towards Mechanistic Systems Biology (Ivano Bertini, K. S. M., Giacomo Parigi, Ed.), Wiley-VCH Verlag GmbH & Co. KGaA.
129. Morozova, L. A., Haynie, D. T., Arico-Muendel, C., Van Dael, H. & Dobson, C. M. (1995). Structural basis of the stability of a lysozyme molten globule. *Nat Struct Biol* 2, 871-5.
130. Keeler, J. (2010). Understanding NMR spectroscopy, Wiley-Blackwell (an imprint of John Wiley & Sons Ltd), Chichester/GB
131. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* 94, 12366-71.
132. Goto, N. K. & Kay, L. E. (2000). New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol* 10, 585-92.
133. Tzakos, A. G., Grace, C. R., Lukavsky, P. J. & Riek, R. (2006). NMR techniques for very large proteins and rnas in solution. *Annu Rev Biophys Biomol Struct* 35, 319-42.
134. Nietlispach, D. (2005). Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. *J Biomol NMR* 31, 161-6.
135. Korzhnev, D. M., Klover, K., Kanelis, V., Tugarinov, V. & Kay, L. E. (2004). Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J Am Chem Soc* 126, 3964-73.

136. Sprangers, R. & Kay, L. E. (2007). Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445, 618-22.
137. Religa, T. L., Sprangers, R. & Kay, L. E. (2010). Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* 328, 98-102.
138. Schanda, P., Kupce, E. & Brutscher, B. (2005). SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* 33, 199-211.
139. Gettins, P. G., Backovic, M. & Peterson, F. C. (2004). Use of NMR to study serpin function. *Methods* 32, 120-9.
140. Nyon, M. P., Kirkpatrick, J., Cabrita, L. D., Christodoulou, J. & Gooptu, B. (2011). ¹H, ¹⁵N and ¹³C backbone resonance assignments of the archetypal serpin α_1 -antitrypsin. *Biomol NMR Assign* 6, 153-6.
141. Mittermaier, A. & Kay, L. E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science* 312, 224-8.
142. Palmer, A. G., 3rd. (2004). NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104, 3623-40.
143. Ishima, R. & Torchia, D. A. (2000). Protein dynamics from NMR. *Nat Struct Biol* 7, 740-3.
144. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277-93.
145. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. (2005).

- The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687-96.
146. Goddard, T. D. K., D. G. ((2000). SPARKY 3.
 147. Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. & Altmann, F. (2006). Comprehensive glyco-proteomic analysis of human α_1 -antitrypsin and its charge isoforms. *Proteomics* 6, 3369-80.
 148. Lomas, D. A., Evans, D. L., Stone, S. R., Chang, W. S. & Carrell, R. W. (1993). Effect of the Z mutation on the physical and inhibitory properties of α_1 -antitrypsin. *Biochemistry* 32, 500-8.
 149. Kirshenbaum, N., Michaelevski, I. & Sharon, M. (2010). Analyzing large protein complexes by structural mass spectrometry. *J Vis Exp*.
 150. Schanda, P. & Brutscher, B. (2006). Hadamard frequency-encoded SOFAST-HMQC for ultrafast two-dimensional protein NMR. *J Magn Reson* 178, 334-9.
 151. Fogh, R., Ionides, J., Ulrich, E., Boucher, W., Vranken, W., Linge, J. P., Habeck, M., Rieping, W., Bhat, T. N., Westbrook, J., Henrick, K., Gilliland, G., Berman, H., Thornton, J., Nilges, M., Markley, J. & Laue, E. (2002). The CCPN project: an interim report on a data model for the NMR community. *Nat Struct Biol* 9, 416-8.
 152. Banerjee, S. K., Holler, E., Hess, G. P. & Rupley, J. A. (1975). Reaction of N-acetylglucosamine oligosaccharides with lysozyme. Temperature, pH, and solvent deuterium isotope effects; equilibrium, steady state, and pre-steady state measurements*. *J Biol Chem* 250, 4355-67.
 153. Zhang, Y.-Z. (1995). Protein and peptide structure and interactions studied by hydrogen exchange and NMR, University of Pennsylvania.

154. Best, R. B. & Vendruscolo, M. (2006). Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of CI2. *Structure* 14, 97-106.
155. Lee, C., Park, S. H., Lee, M. Y. & Yu, M. H. (2000). Regulation of protein function by native metastability. *Proc Natl Acad Sci U S A* 97, 7727-31.
156. Knaupp, A. S. & Bottomley, S. P. (2009). Serpin polymerization and its role in disease--the molecular basis of α_1 -antitrypsin deficiency. *IUBMB Life* 61, 1-5.
157. Yu, M. H., Lee, K. N. & Kim, J. (1995). The Z type variation of human α_1 -antitrypsin causes a protein folding defect. *Nat Struct Biol* 2, 363-7.
158. Lomas, D. A. (1996). New insights into the structural basis of α_1 -antitrypsin deficiency. *QJM* 89, 807-12.
159. Shental-Bechor, D. & Levy, Y. (2009). Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Curr Opin Struct Biol* 19, 524-33.
160. Stein, P. E. & Carrell, R. W. (1995). What do dysfunctional serpins tell us about molecular mobility and disease? *Nat Struct Biol* 2, 96-113.
161. Thirumalai, D., O'Brien, E. P., Morrison, G. & Hyeon, C. (2010). Theoretical perspectives on protein folding. *Annu Rev Biophys* 39, 159-83.
162. Lim, W. K., Rosgen, J. & Englander, S. W. (2009). Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc Natl Acad Sci U S A* 106, 2595-600.

163. Clark, E. D. B. (1998). Refolding of recombinant proteins. *Curr Opin Biotechnol* 9, 157-63.
164. Milardi, D., La Rosa, C. & Grasso, D. (1994). Extended theoretical analysis of irreversible protein thermal unfolding. *Biophys Chem* 52, 183-9.
165. Bruch, M., Weiss, V. & Engel, J. (1988). Plasma serine proteinase inhibitors (serpins) exhibit major conformational changes and a large increase in conformational stability upon cleavage at their reactive sites. *J Biol Chem* 263, 16626-30.
166. Knaupp, A. S., Keleher, S., Yang, L., Dai, W., Bottomley, S. P. & Pearce, M. C. (2013). The roles of helix I and strand 5A in the folding, function and misfolding of α_1 -antitrypsin. *PLoS One* 8, e54766.
167. Pearce, M. C., Morton, C. J., Feil, S. C., Hansen, G., Adams, J. J., Parker, M. W. & Bottomley, S. P. (2008). Preventing serpin aggregation: the molecular mechanism of citrate action upon antitrypsin unfolding. *Protein Sci* 17, 2127-33.
168. Kim, J., Lee, K. N., Yi, G. S. & Yu, M. H. (1995). A thermostable mutation located at the hydrophobic core of α_1 -antitrypsin suppresses the folding defect of the Z-type variant. *J Biol Chem* 270, 8597-601.
169. Parfrey, H., Mahadeva, R., Ravenhill, N. A., Zhou, A., Dafforn, T. R., Foreman, R. C. & Lomas, D. A. (2003). Targeting a surface cavity of α_1 -antitrypsin to prevent conformational disease. *J Biol Chem* 278, 33060-6.
170. Gooptu, B., Miranda, E., Nobeli, I., Mallya, M., Purkiss, A., Brown, S. C., Summers, C., Phillips, R. L., Lomas, D. A. & Barrett, T. E. (2009). Crystallographic and cellular characterisation of two mechanisms

- stabilising the native fold of α_1 -antitrypsin: implications for disease and drug design. *J Mol Biol* 387, 857-68.
171. Knaupp, A. S., Levina, V., Robertson, A. L., Pearce, M. C. & Bottomley, S. P. (2010). Kinetic instability of the serpin Z α_1 -antitrypsin promotes aggregation. *J Mol Biol* 396, 375-83.
 172. Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat Struct Biol* 4, 630-4.
 173. Nyon, M. P., Segu, L., Cabrita, L. D., Levy, G. R., Kirkpatrick, J., Roussel, B. D., Patschull, A. O., Barrett, T. E., Ekeowa, U. I., Kerr, R., Waudby, C. A., Kalsheker, N., Hill, M., Thalassinou, K., Lomas, D. A., Christodoulou, J. & Gooptu, B. (2012). Structural dynamics associated with intermediate formation in an archetypal conformational disease. *Structure* 20, 504-12.
 174. Belorgey, D., Hagglof, P., Karlsson-Li, S. & Lomas, D. A. (2007). Protein misfolding and the serpinopathies. *Prion* 1, 15-20.
 175. Dalby, P. A., Clarke, J., Johnson, C. M. & Fersht, A. R. (1998). Folding intermediates of wild-type and mutants of barnase. II. Correlation of changes in equilibrium amide exchange kinetics with the population of the folding intermediate. *J Mol Biol* 276, 647-56.
 176. Baek, J. H., Yang, W. S., Lee, C. & Yu, M. H. (2009). Functional unfolding of α_1 -antitrypsin probed by hydrogen-deuterium exchange coupled with mass spectrometry. *Mol Cell Proteomics* 8, 1072-81.
 177. Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T. & Dobson, C. M. (1993). Detection of transient protein folding populations by mass spectrometry. *Science* 262, 896-900.

178. Yang, H. & Smith, D. L. (1997). Kinetics of cytochrome c folding examined by hydrogen exchange and mass spectrometry. *Biochemistry* 36, 14992-9.
179. Clarke, J. & Itzhaki, L. S. (1998). Hydrogen exchange and protein folding. *Curr Opin Struct Biol* 8, 112-8.
180. Linderstrøm-Lang, K. U. (1958). Deuterium exchange and protein structure. Neu-Berger A. Symposium on Protein Structure.
181. Englander, S. W., Englander, J. J., McKinnie, R. E., Ackers, G. K., Turner, G. J., Westrick, J. A. & Gill, S. J. (1992). Hydrogen exchange measurement of the free energy of structural and allosteric change in hemoglobin. *Science* 256, 1684-7.
182. Englander, S. W. (1975). Measurement of structural and free energy changes in hemoglobin by hydrogen exchange methods. *Ann N Y Acad Sci* 244, 10-27.
183. Wand, A. J., Roder, H. & Englander, S. W. (1986). Two-dimensional ¹H NMR studies of cytochrome c: hydrogen exchange in the N-terminal helix. *Biochemistry* 25, 1107-14.
184. Englander, S. W., Downer, N. W. & Teitelbaum, H. (1972). Hydrogen exchange. *Annu Rev Biochem* 41, 903-24.
185. Englander, J. J., Del Mar, C., Li, W., Englander, S. W., Kim, J. S., Stranz, D. D., Hamuro, Y. & Woods, V. L., Jr. (2003). Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc Natl Acad Sci U S A* 100, 7057-62.
186. Clarke, A. R. & Waltho, J. P. (1997). Protein folding and intermediates. *Curr Opin Biotechnol* 8, 400-10.

187. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* 29, 213-38.
188. Linderstrom. (1958). Ed. Symposium on Protein Structure.
189. Miranda, E., Perez, J., Ekeowa, U. I., Hadzic, N., Kalsheker, N., Gooptu, B., Portmann, B., Belorgey, D., Hill, M., Chambers, S., Teckman, J., Alexander, G. J., Marciniak, S. J. & Lomas, D. A. (2010). A novel monoclonal antibody to characterize pathogenic polymers in liver disease associated with α_1 -antitrypsin deficiency. *Hepatology* 52, 1078-88.
190. Mahadeva, R., Chang, W. S., Dafforn, T. R., Oakley, D. J., Foreman, R. C., Calvin, J., Wight, D. G. & Lomas, D. A. (1999). Heteropolymerization of S, I, and Z α_1 -antitrypsin and liver cirrhosis. *J Clin Invest* 103, 999-1006.
191. Fra, A. M., Gooptu, B., Ferrarotti, I., Miranda, E., Scabini, R., Ronzoni, R., Benini, F., Corda, L., Medicina, D., Luisetti, M. & Schiaffonati, L. (2012). Three new α_1 -antitrypsin deficiency variants help to define a C-terminal region regulating conformational change and polymerization. *PLoS One* 7, e38405.
192. Cabrita, L. D., Whisstock, J. C. & Bottomley, S. P. (2002). Probing the role of the F-helix in serpin stability through a single tryptophan substitution. *Biochemistry* 41, 4575-81.
193. Cabrita, L. D., Dai, W. & Bottomley, S. P. (2004). Different conformational changes within the F-helix occur during serpin folding, polymerization, and proteinase inhibition. *Biochemistry* 43, 9834-9.

194. Pearce, M. C., Cabrita, L. D., Rubin, H., Gore, M. G. & Bottomley, S. P. (2004). Identification of residual structure within denatured antichymotrypsin: implications for serpin folding and misfolding. *Biochem Biophys Res Commun* 324, 729-35.
195. Elliott, P. R., Lomas, D. A., Carrell, R. W. & Abrahams, J. P. (1996). Inhibitory conformation of the reactive loop of α_1 -antitrypsin. *Nat Struct Biol* 3, 676-81.
196. Janciauskiene, S., Eriksson, S., Callea, F., Mallya, M., Zhou, A., Seyama, K., Hata, S. & Lomas, D. A. (2004). Differential detection of PAS-positive inclusions formed by the Z, Siiyama, and Mmalton variants of α_1 -antitrypsin. *Hepatology* 40, 1203-10.
197. Carrell, R. W. & Lomas, D. A. (1997). Conformational disease. *Lancet* 350, 134-8.
198. Carrell, R. W., Lomas, D. A., Sidhar, S. & Foreman, R. (1996). α_1 -antitrypsin deficiency. A conformational disease. *Chest* 110, 243S-247S.
199. Perry, D. J., Marshall, C., Borg, J. Y., Tait, R. C., Daly, M. E., Walker, I. D. & Carrell, R. W. (1995). Two novel antithrombin variants, Asn187Asp and Asn187Lys, indicate a functional role for asparagine 187. *Blood Coagul Fibrinolysis* 6, 51-4.
200. Bruce, D., Perry, D. J., Borg, J. Y., Carrell, R. W. & Wardell, M. R. (1994). Thromboembolic disease due to thermolabile conformational changes of antithrombin Rouen-VI (187 Asn-->Asp). *J Clin Invest* 94, 2265-74.
201. Gettins, P. G. (2002). The F-helix of serpins plays an essential, active role in the proteinase inhibition mechanism. *FEBS Lett* 523, 2-6.

202. Gooptu, B., Ekeowa, U. I. & Lomas, D. A. (2009). Mechanisms of emphysema in α_1 -antitrypsin deficiency: molecular and cellular insights. *Eur Respir J* 34, 475-88.
203. Sharp, A. M., Stein, P. E., Pannu, N. S., Carrell, R. W., Berkenpas, M. B., Ginsburg, D., Lawrence, D. A. & Read, R. J. (1999). The active conformation of plasminogen activator inhibitor 1, a target for drugs to control fibrinolysis and cell adhesion. *Structure* 7, 111-8.
204. Dobson, C. M. (2001). The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci* 356, 133-45.
205. Dobson, C. M. (1995). Finding the right fold. *Nat Struct Biol* 2, 513-7.
206. Dobson, C. M. (2002). Getting out of shape. *Nature* 418, 729-30.
207. Selkoe, D. J. (2003). Folding proteins in fatal ways. *Nature* 426, 900-4.
208. Lomas, D. A. & Mahadeva, R. (2002). α_1 -antitrypsin polymerization and the serpinopathies: pathobiology and prospects for therapy. *J Clin Invest* 110, 1585-90.
209. Davis, R. L., Shrimpton, A. E., Carrell, R. W., Lomas, D. A., Gerhard, L., Baumann, B., Lawrence, D. A., Yepes, M., Kim, T. S., Ghetti, B., Piccardo, P., Takao, M., Lachawan, F., Muenke, M., Sifers, R. N., Bradshaw, C. B., Kent, P. F., Collins, G. H., Larocca, D. & Holohan, P. D. (2002). Association between conformational mutations in neuroserpin and onset and severity of dementia. *Lancet* 359, 2242-7.
210. Huntington, J. A., Sendall, T. J. & Yamasaki, M. (2009). New insight into serpin polymerization and aggregation. *Prion* 3, 12-4.

211. Mushero, N. & Gershenson, A. (2011). Determining serpin conformational distributions with single molecule fluorescence. *Methods Enzymol* 501, 351-77.
212. Tsutsui, Y., Sarkar, A. & Wintrode, P. L. (2011). Probing serpin conformational change using mass spectrometry and related methods. *Methods Enzymol* 501, 325-50.
213. Bird, J. C. W. P. L., Ed. (2011). *Methods in enzymology*. Vol. 501. Serpin structure and evolution Edited by Simon, J. N. A. M. I.: Elsevier.
214. Chung, E. W., Nettleton, E. J., Morgan, C. J., Gross, M., Miranker, A., Radford, S. E., Dobson, C. M. & Robinson, C. V. (1997). Hydrogen exchange properties of proteins in native and denatured states monitored by mass spectrometry and NMR. *Protein Sci* 6, 1316-24.
215. Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. (2003). Protein hydrogen exchange mechanism: local fluctuations. *Protein Sci* 12, 153-60.
216. Hansen, D. F., Vallurupalli, P. & Kay, L. E. (2009). Measurement of methyl group motional parameters of invisible, excited protein states by NMR spectroscopy. *J Am Chem Soc* 131, 12745-54.
217. Tugarinov, V., Sprangers, R. & Kay, L. E. (2007). Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *J Am Chem Soc* 129, 1743-50.
218. Lundstrom, P., Lin, H. & Kay, L. E. (2009). Measuring ¹³Cβ chemical shifts of invisible excited states in proteins by relaxation dispersion NMR spectroscopy. *J Biomol NMR* 44, 139-55.
219. Korzhnev, D. M., Religa, T. L., Lundstrom, P., Fersht, A. R. & Kay, L. E. (2007). The folding pathway of an FF domain: characterization of an

- on-pathway intermediate state under folding conditions by (^{15}N) , $(^{13}\text{C}\alpha)$ and (^{13}C) -methyl relaxation dispersion and $(^1\text{H})/(^2\text{H})$ -exchange NMR spectroscopy. *J Mol Biol* 372, 497-512.
220. Hore, P. J., Jones, J. A. & Wimperis, S. (2000). *NMR: The toolkit*, Oxford University Press, Oxford.
 221. Abraham, R. J., Fisher, J. & Loftus, P. . (1988). *Introduction to NMR spectroscopy*, John Wiley and Sons, Inc., Chichester.
 222. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1997). Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* 36, 2390-402.
 223. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28, 8972-9.
 224. Lakomek, N. A., Ying, J. & Bax, A. (2012). Measurement of $(^1)(^{15})\text{N}$ relaxation rates in perdeuterated proteins by TROSY-based methods. *J Biomol NMR* 53, 209-21.
 225. Lipari, G. S., A. . (1982). relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.*
 226. Lipari, G. S., A. . (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.*
 227. Mandel, A. M., Akke, M. & Palmer, A. G., 3rd. (1995). Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* 246, 144-63.

228. Korzhnev, D. M., Neudecker, P., Mittermaier, A., Orekhov, V. Y. & Kay, L. E. (2005). Multiple-site exchange in proteins studied with a suite of six NMR relaxation dispersion experiments: an application to the folding of a Fyn SH3 domain mutant. *J Am Chem Soc* 127, 15602-11.
229. Mulder, F. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. (2001). Studying excited states of proteins by NMR spectroscopy. *Nat Struct Biol* 8, 932-5.
230. Kovtun, O., Mureev, S., Jung, W., Kubala, M. H., Johnston, W. & Alexandrov, K. (2011). Leishmania cell-free protein expression system. *Methods* 55, 58-64.